Signal Processing and Biomedical Analysis of Data Collected from Wearable Devices in order to Evaluate Individualized Circadian Rhythm and Reliably Extract Cardiac Vital Signs from Ambulatory Electrocardiograms

Gregory Laredo

Department of Biological & Biomedical Engineering

McGill University, Montreal

December 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Engineering

Acknowledgments

With sincere gratitude, I would like to thank my supervisor, Georgios D. Mitsis for all his training, guidance and support throughout my degree and for allowing me to pursue my research aspirations.

I would like to thank the entire Biosignals & Systems Analysis Lab team during my research tenure, specifically Arna Gosh, Michalis Kassinopoulos, Kyriaki Kostoglou, Sara Larivière, Dylan Mann-Krzisnik, Prokopis Prokopiou and Alba Xifra-Porxas. I learned tremendously from them and am thankful for the amazing work environment we had together.

I would like to acknowledge Thinkinetics Inc. and the Elastic Care Inc. for collaborating with my supervisor and I on the first and second parts of my research, respectively. Their vision and application of research were extremely valuable.

With reverence, I would like to express my sincere appreciation to my parents, Max H. and Sigy Laredo, for all their support and encouragement throughout this degree and all that came before. They always motivated me and provided me with the resources necessary to succeed in my education. I aspire to do the same for my children.

To my beloved wife, Dana, I thank you for your constant motivation and support at each step of the way. If it was not for you, I would not be where I am today. I hope to continue making you proud. And to my dear children, Neomi and Mikayla, you are my inspiration for this achievement and all others to come.

Gratitude to the rest of my family and friends who contributed directly or indirectly towards helping me complete my thesis.

Contributions of Authors

The body of work developed in this thesis would not have been made possible without the constant feedback and direction from my supervisor, **Georgios D. Mitsis**, specifically:

For Research Part I: Provided advice for estimating and analyzing 24-hour fluctuations in HR recorded from a smartwatch over multiple days with intervals of data interruptions, and helped with interpreting the results.

For Research Part II: Suggested training and Random Forest Classifier for proposing a novel signal quality index for ambulatory ECG data based on features brought down in literature, helped interpret results, and provided advice on how to statistically evaluate results.

My labmate, **Kyriaki Kostoglou**, greatly contributed in research analysis, specifically:

For Research Part I: Provided MATLAB code for signal smoothing operations and provided advice on parameterization models and evaluation criteria.

For Research Part II: Provided MATLAB code for training and testing a Random Forest Classifier she co-developed and helped with initializing variables and interpreting results.

Abstract

Background: Wearable devices capable of monitoring physiological data is a growing field of interest with several applications on how we can analyze and use the data to benefit our everyday lives. Despite the leaps in technological advancements, there still remains several challenges with regards to collecting data over multiple days for longitudinal studies, such as circadian rhythm analysis, as well as ambulatory environment artifacts corrupting signals and compromising the monitoring required for real-time information and decision making, such as estimating cardiac vital signs of an outpatient.

Purpose: The main objective of this thesis is to reliably collect and interpret physiological data collected from wearable devices and develop algorithms that creatively provide this information to the users in a way that can easily be employed by a wearable device.

Methods: Two parts comprise this thesis: (Part I) based on heart rate (HR) data recorded from wearable devices, the endogenous, diurnal fluctuations of an individual's HR was investigated in the free-living environment of everyday life, circadian HR profile with key anchors and its relationship with sleep-wake schedule identified, and an individualized model characterizing circadian fluctuations of HR proposed; and (Part II) a real-time algorithm for processing ambulatory electrocardiograms (ECG) was developed with a modified Pan-Tompkins QRS detection algorithm incorporated to detect heart beats, which were then used to estimate HR and heart rate variability with every second of recording accompanied by a novel signal quality index (SQI) quantifying how reliable the R-peak derived cardiac vital signs are.

Results: (Part I): The endogenous, diurnal fluctuations of HR was successfully identified from a smartwatch with HR monitoring capabilities, key anchors were found to very closely relate to the individual's sleep-wake schedule, as reported in literature, and a Fourier Series parameterization model was used to characterize the individual's circadian HR profile. (Part II): The R-peak detection algorithm developed yielded very high performance on the standard ambulatory ECG database and the SQI proposed leveraged a Random Forest classifier analysis of important features and provided strong evidence to capture information relating with what degree the cardiac vital signs outputted should be considered reliable.

Abstrait (French)

Contexte: Les dispositifs portables capables de surveiller des données physiologiques constituent un domaine d'intérêt croissant avec de nombreuses applications sur la façon dont nous pouvons analyser et utiliser les données pour améliorer notre vie quotidienne. Malgré les avancées technologiques, la collecte de données sur plusieurs jours pour des études longitudinales telles que l'analyse du rythme circadien, ainsi que des artefacts de l'environnement ambulatoire qui corrompent les signaux et compromettent la surveillance en temps réel et prise de décision, telles que l'estimation des signes vitaux cardiaques d'un patient ambulatoire.

Objectif: L'objectif principal de cette thèse est de collecter et d'interpréter de manière fiable les données physiologiques recueillies à partir de dispositifs portables et de développer des algorithmes qui fournissent de manière créative ces informations aux utilisateurs dans une façon qui peuvent facilement être utilisées par un dispositif portable.

Méthodes: Cette thèse est composée de deux parties: (Partie I) basée sur les données de fréquence cardiaque (HR) enregistrées à partir de dispositifs portables, les fluctuations endogènes et diurnes de la HR d'une personne ont été étudiées dans le cadre de vie libre de la vie quotidienne, le profil de la FC circadien avec les points d'ancrage clés et leur relation avec le calendrier veille-sommeil ont été identifiés, et un modèle individualisé caractérisant les fluctuations circadiennes de la FC proposé; et (Partie II) un algorithme en temps réel pour le traitement des électrocardiogrammes ambulatoires (ECG) a été développé avec un modifié Pan-Tompkins algorithme de détection QRS incorporé pour détecter les battements cardiaques, qui ont ensuite été utilisés pour estimer la variabilité de la fréquence cardiaque et de la fréquence cardiaque à chaque seconde d'enregistrement accompagné d'un nouvel indice de qualité du signal (SQI) quantifiant la fiabilité des signes vitaux cardiaques dérivés du pic-R.

Résultats: (Partie I): Les fluctuations endogènes et diurnes de la FC ont été identifiées avec succès grâce à une smartwatch dotée de capacités de surveillance de la FC. modèle de paramétrisation a été utilisé pour caractériser le profil de fréquence cardiaque circadien de la personne. (Partie II): L'algorithme de détection des pics R développé a donné de très hautes performances sur la base de données ECG ambulatoire standard et, en utilisant une Forêt Aléotoire analyse des characteristic important, le SQI proposé a fourni des preuves solides pour capturer des informations relatives au degré de fiabilité des signes vitaux cardiaques produits.

Original Contributions

Research Part I:

The diurnal, endogenous fluctuations of an individual's HR from HR data recorded from commercially available wearable devices was investigated over multiple days in unrestricted, free-living conditions. Through creative circadian smoothing operations, key anchors commonly reported in controlled studies were confirmed and an innovative method for displaying the individual's probability of sleep-wake cycle was integrated to provide a new layer of information for the first time in uncontrolled circadian rhythm studies.

Research Part II:

In this study, a real-time ambulatory ECG processing algorithm detecting R-peaks and outputting second-by-second heart rate and heart rate variability measurements, respectively, was developed. A unique method for defining a weighted combination of features and decision rules was proposed to accompany each output in order to relay a measure of how reliable the R-peak derived cardiac vital signs are (real number ranging from 0 to 1). This was achieved from a novel analysis of feature importance derived from a Random Forest classifier trained on a noisy dataset of increasing signal-to-noise ratio with passed or failed R-peak detection performance as the binary classification labels.

Table of Contents

List of Abbreviations

Avg.	average
bpm	beats per minute
CBT	core body temperature
CVD	cardiovascular diseases
ECG	electrocardiogram
FDA	Food and Drug Administration
FN	false negative
FP	false positive
h	hour
HPF	high pass filter
HR	heart rate
HRV	heart rate variability
LPF	low pass filter
MIT-BIH	Massachusetts Institute of Technology - Beth Israel Hospital databases
MLII	modified limb lead II
MPE	mean percent error
ms	millisecond
mV	millivolts
NST	Noise Stress Test database
OOB	out-of-bag
Pp	positive predictivity
PTT	pulse transit time
RF	Random Forest
RMSE	root-mean squared-error
RQ	respiratory quotient
SCN	suprachiasmatic nucleus
Se	sensitivity
sec	seconds
SE	standard error
SQI	signal quality index
Std.	standard deviation
SVM	Support Vector Machine
TP	true positive
USW	ultradian sleep-wake cycle

* * *

Chapter 1
Introduction to Thesis Research

1 Introduction

1.1 General Introduction

The field of wearable devices is growing rapidly and unlocks exciting applications. In 2018, 177.2 million wearable devices were shipped, up 27.5% from the previous year [1]. Analysts project the wearable device market will be worth USD 25 billion by the end of 2019 and grow at a compounded annual growth rate of 17.7% over the next 5 years [2][3]. The different segments of wearables are separated into consumer, defense, healthcare, industrial and other applications, ordered by increasing market share [4]. The major companies paving the road to bring wearable devices to the market are Apple (market leader), Samsung, LG, Fitbit, Xiaomi, Nike, Garmin, Huawei, Abbot and Medtronic amongst others [5]. Several consumer wearables with health and fitness applications incorporate sensors which monitor biological signals, and of course medical-grade healthcare wearables do the same. Currently, there are consumer wearables that monitor heart rate (HR), heart rate variability (HRV), electrocardiogram (ECG), blood pleasure, breathing rate, electromyogram, sweat analytes, biomechanics movement, and even cognitive function [6]. Wearables are predominantly in the form of smartwatches (approximately 80% market share), however other form factors such as bands, garments, patches, earwear, eyewear and rings with embedded sensors exist as well [7]. As more sensors, functionalities and valuable information are developed – which I argue need to be related to health and wellness – wearables can quickly cross the line from being an accessary to being a necessity, similarly to how today a smartphone is practically a necessity in the modern developed world.

In order for wearables to win the case of being an essential piece of technology, one area requiring continuous efforts is signal integrity and physiological monitoring reliability. Recording biological signals in the ambulatory environment inherently understands that the user can constantly be in motion which can be expected to introduce a suite of artifacts corrupting the signal. This potentially means the health/medical parameter intending to extract from the signal becomes more challenging and unreliable if no signal processing or intelligent algorithms are implemented

to mitigate the noise effect. Furthermore, the ambulatory environment is unpredictable and introduces unknown factors effecting physiological parameters which cannot always be detected, interpreted and/or accounted for in the output. Hence, this is another form of potentially unreliable information about biological states provided to the user. Together, these are two ways misinformation can be provided to the user that at best deters adaptation from the general consumer, however more seriously can lead to misdiagnoses when considering medical wearable devices worn by patients relying on accurate, real-time data for dear life.

1.2 Research Objectives

The broad objective of my thesis research was to reliably collect and interpret physiological signals recorded from wearable devices. More explicitly, this involved familiarizing myself with the experimental process and raw format of collecting data from wearable devices, employing signal processing techniques in order to mitigate noise factors and best retrieve underlying physiological signals, and developing algorithms which extracts reliable biological information about the subject wearing the device and then outputs them in a easy to understand way that facilitates analysis and interpretation. Furthermore, all operations were done with a design focus that would allow for real-time implementation on devices in the field. My thesis research project was split into two parts:

Part I: Identification of individualized circadian fluctuations under free-living conditions using HR data collected from wearable devices.

Part II: Real-time ambulatory ECG processing and novel signal quality index (SQI) for R-peak-derived cardiac vital signs.

In Part I, the specific objectives were to first evaluate whether circadian fluctuations of an individual's HR can be monitored in free-living conditions from a smartwatch with HR-monitoring capability and then propose a model for quantifying HR's circadian fluctuation along with its dependency on external factors such as sleep-wake schedule.

In Part II, the main objective was to develop a real-time algorithm for processing ambulatory ECG data collected from a medical-grade remote patient monitoring device. The outputs of interest were cardiac vital signs, HR and HRV, as well as a novel SQI which quantifies the reliability of the underlying ECG segment and the beat detections used to derive the aforementioned cardiac vital signs.

Chapter 2

Research Part I

2 Identification of individualized circadian fluctuations under free-living conditions using HR data collected from wearable devices

2.1 Introduction

Over the past several decades, increasing scientific evidence of an internal biological clock in humans have emerged. Referred to as the circadian (Latin for "around a day") rhythm, nearly all living organisms demonstrate endogenous oscillations in physiological, metabolic and/or behavioral functions over the roughly 24 hours of each day [8]. Researchers have identified this in human beings on the cellular level all the way up to the organ systems level through monitoring a wide array of physiological parameters over at least one uninterrupted day; for example: transcription genes, protein and ion concentrations, hormone secretion and sugar levels, vital signs, such as HR, respiration rate and core body temperature (CBT), and brain activity [9]-[12]. Each have been found to display evident and characteristic circadian rhythmicity under controlled environments. The main factors which have been found to entrain as well as disrupt circadian rhythms, referred to as *zeitgebers* in the field of chronobiology, are exposure to light, sleep-wake schedule, active-rest intervals, eating-fasting distribution, temperature, social interactions and medications [13]. As a result, it is common practice to identify and study disruptions of circadian rhythms in humans, respectively, under highly controlled laboratory settings where all *zeitgebers* can be monitored and regulated and then individually disrupted depending on the aim of the study.

More recently, a growing number of studies have linked the desynchronization of people's activities with their internal biological clock to several of the most common diseases and disorders: obesity, diabetes, depression, cardiovascular diseases, cancer, and more [14]-[18]. Similarly, studies on the treatment and therapy of diseases and disorders have recently started considering the chronobiologic implications on efficacy and prognosis; for example, chronochemotherapy for cancer treatment considers the pharmacokinetics and pharmacodynamics of chemotherapeutic

drugs at certain points in the biological clock as well as proliferation of healthy cells and activity of DNA repair enzymes [17].

Advancements in wearable technology now provide us the ability to record some of these physiological parameters identified to display circadian fluctuations, this time however, outside the laboratory. Thus, monitoring whether or not endogenous fluctuations are evident and significant in everyday life and understanding which and how specific activities/events contribute towards synchronization or desynchronization is for the first time at arm's reach. With all the studies and evidence confirming human beings' biological functioning and dependency on a circadian rhythm, there still remains a large gap between what is concluded by scientists in laboratories and how it can be used to inform actionable insights in the free-living environment of everyday life. Through the use of wearable technology, signal processing and data science, this study attempts to narrow this gap and help lay the groundwork to a healthier lifestyle in-synched with individualized circadian rhythm.

2.1.1 Research Objectives and Scope

The overarching objective of this research is to test whether the diurnal, endogenous fluctuations of a single physiological parameter – HR – can be identified from a smartwatch currently available on the consumer market. The smartwatch must of course have HR monitoring capabilities and will then be validated as a reliable signal that records cardiac beats. Thereafter, research continued into developing a model which characterizes individual-specific circadian fluctuations of HR and its relationship with sleep-wake schedule.

The scope of this study involved monitoring HR from one individual subject over multiple weeks under complete free-living environment conditions. What this means is that no fixed schedule or restriction of any sort was imposed on the subject; they were at liberty to choose their sleep-wake schedule, eating-fasting periods and active-rest intervals, all which are known external factors contributing to non-endogenous changes in HR. Therefore, the effects of external factors was taken into consideration during signal processing and interpretation.

2.2 Background

What is theorized to have originated from Earth's rotation about its axis in the Solar System and the consequential day and night cycles established with our Sun, today almost all lifeform –

bacteria, fungi, plants, animals – possesses a timekeeping system which regulates organism functions with a roughly 24 hour period [8]. The primary environmental factor resposnible for entraining this circadian rhythm is sunlight exposure or light and dark stimuli. The most high level circadian regulation known to human beings may very well be our sleep and wake schedule; however, there are many other physiological processes similarly regulated on this time scale, including body temperatures, hormone release, feeding behaviour, drug metabolism, glucose homeostatis and cell-cycle progression [19]. Evidents suggests that a master internal clock exists and is responsible for all these physiological circadian rhythms through neuronal and hormonal regulation. This master clock, also referred to as the central circadian pacemaker, is locatted in mammals in the suprachiasmatic nucleus (SCN) of the hypothalumus. **Figure 1** illustrates the general anatomical organization of the hypothalomus and some of the other components involved in the central circadian pacemaker. The SCN comprises approximately 20,000 neurons organized in a bilateral nuclei bed and is situated directly above the optic chiasm. Specialized photosensitive retinal ganglion cells receive photic information about the daily light cycle from the retina and sends it to the SCN via the retinohypothalmic tract. From there, a cascade of neural pathways and hormone secreation signaling, originating from the SCN, control the wide array of physiological functions known (and yet to be known) to be regulated on a circadian rhythm. [19]

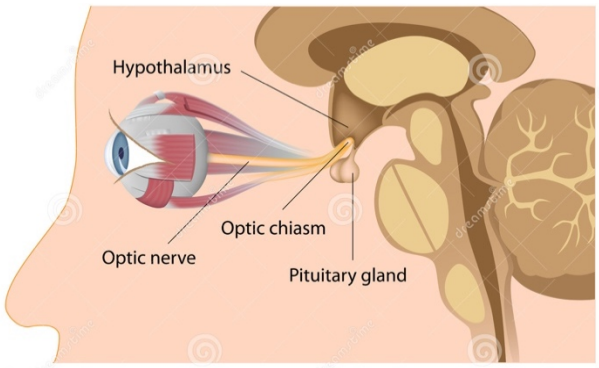

Figure 1: General anatomical organization of components involved in the central circadian pacemaker situated in the suprachiasmatic nucleus (SCN) of the hypothalamus [20].

Endogenous fluctuations of human beings' HR is but one physiological parameter which has been studied to follow a circadian rhythm. Researchers have been very interested in monitoring vital signs such as HR, respiration rate and body temperatures, as well as hormone concentration levels (melatonin and cortisol predominantly) in order to study circadian processes and understand

the synchronization and desynchronization of the master clock. **Figure 2** below, prepared by Dr. Diane Boivin, director of the Centre of Study and Treatment of Circadian Rhythms, Douglas Research Centre, presents a generalized circadian profile of some of the main physiological parameters known to exhibit endogenous circadian fluctuations. When it comes to human's cyclic sleep-wake schedule, Dr. Boivin explains so eloquently how there are two processes which interact with each other in order to keep healthy adults awake for about 16 hours and asleep for about 8 hours each day; these are the homeostatic process and the circadian process. The homeostatic process drives the propensity for one to sleep the longer they stay awake, and the circadian process drives favourable physiological conditions for sleep at different times of the day. [21]

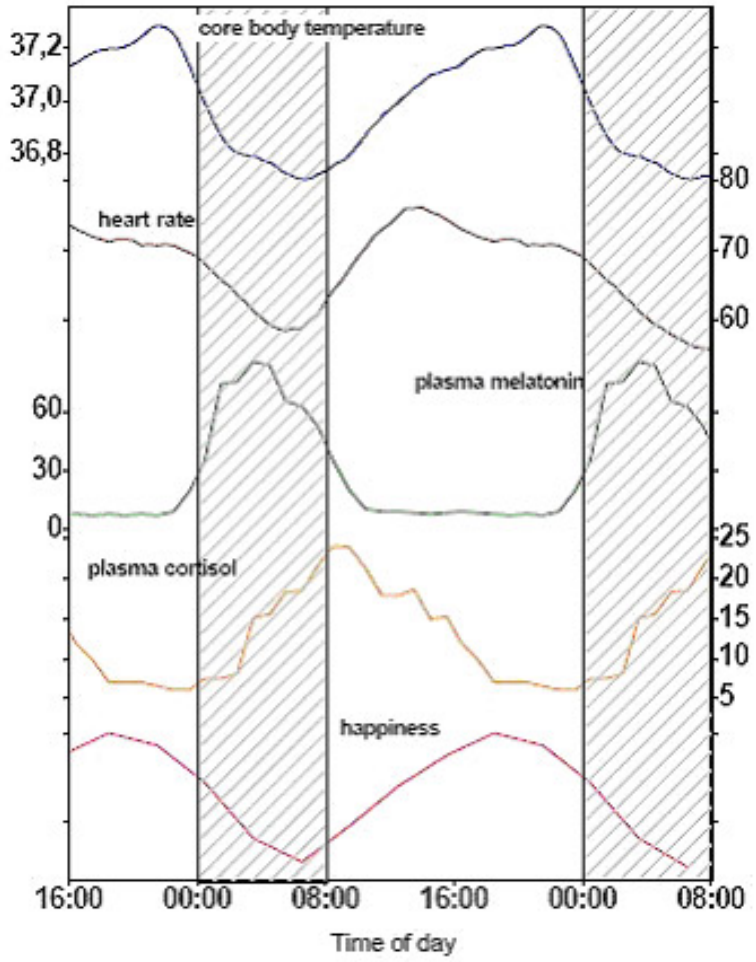

Figure 2: Generalized circadian profile of core body temperature, heart rate, plasma melatonin, plasma cortisol and happiness. [21]

In 1994, a novel study by Krauchi and Wirz-Justice on the circadian rhythm of heat production, HR, respiratory quotient (RQ) and body temperatures was published [11]. Their study consisted of recording these vital signs, as well as urine data for hormone and ion levels, in seven healthy adult males over 34 hours of constant bed rest with no sleep allowed and regular isocaloric food and fluid intake. Because of this constant routine protocol, referred to as unmasking conditions, the researchers hypothesized that if vital signs fluctuated it would be as a result of an

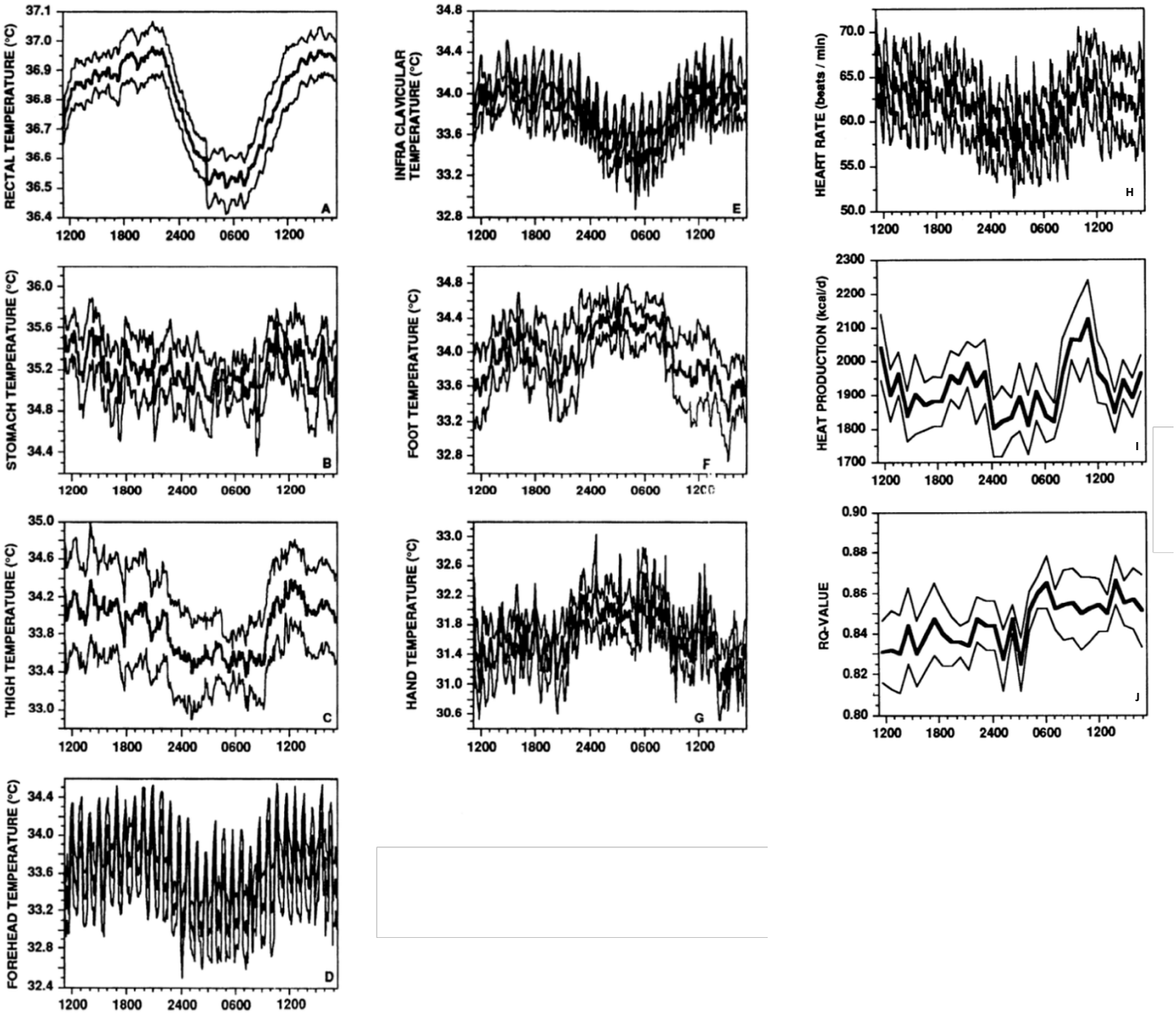

Figure 3: Circadian rhythm over 30.5 hours in rectal **(A)** and skin **(B-G)** temperatures, heart rate **(H)**, heat production **(I)** and respiratory quotient (RQ) value **(J)**. Data points are means +/- SE of 7 men at 5 minute intervals for all signals except heat production and RQ-value at 4 hour intervals. Thick line: mean; thin line: +/- 1 SE; abscissa: time of day in hours. [11]

endogenous circadian component. The conclusion of their study showed that heat production, HR and rectal and distal skin temperatures demonstrate a significant circadian rhythm, however stomach temperature and RQ do not. **Figure 3** above presents the summary recordings of these parameters over the 34 hours as mean ± standard error (SE) for the seven subjects tested. It is evident in these results that HR, amongst other parameters, display significant circadian fluctuation profile. Krauchi and Wirz-Justice further quantified the minimum and maximum values, respectively, of each vital sign as a mean ± SE for the seven subjects and the one-hour range at which they occurred at. They were then able establish which physiological parameters are phase locked or opposite in phase. For HR, they derived that the minimum HR value was 58.22 ± 3.29 bpm occurring over the hour 0400-0500 and the maximum HR value was 64.64 ± 3.85 bpm occurring over the hour 1100-1200. In circadian rhythm studies, the CBT nadir, usually estimated from rectal temperature, in a 24-hour cycle is commonly used as a reference and assigned a circadian phase equal to 0. Units of circadian phase can either be expressed in hours, thus ranging between 0 and 24 hours and setting CBT nadir equal to 0 hour, or in units of degrees after segmenting the 24 hours in a day into 360 degrees and similarly setting the CBT nadir to 0 degrees.

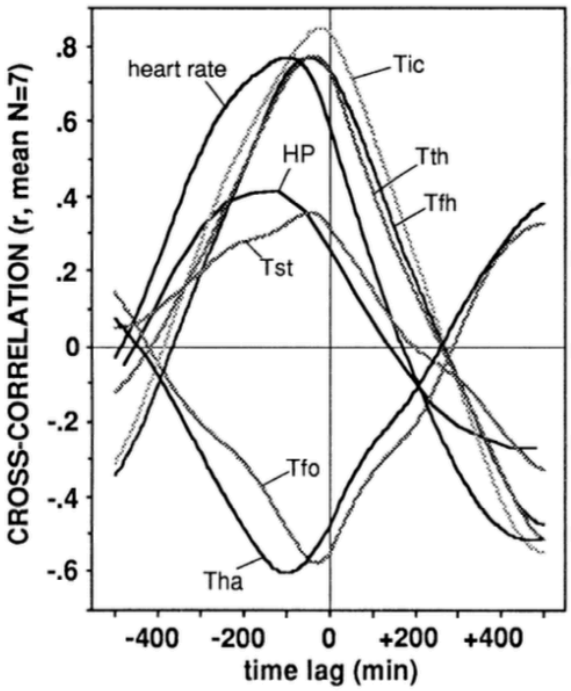

Figure 4: Mean cross-correlation coefficients (after Fisher's Z-transformation), calculated separately for the 7 subjects, between rectal temperature and heart rate, heat production (HP) and all skin temperatures (hand: Tha; foot: Tfo; stomach: Tst; midthigh: Tth; infraclavicular: Tic; and forehead: Tfh). [11]

Figure 4 presents the cross-correlation coefficient analysis of each parameter with rectal skin temperature, which was used to determine phase relationships. From these results, this study concluded that HR is phase locked with rectal temperature as well as with heat production and proximal skin temperatures (infraclavicular, thigh and forehead) however opposite in phase with distal skin temperatures (hands and feet). From this study, the authors argue that despite many studies of the time claiming day-night variations in HR is entirely exogenous and only related to activity level, this study's unmasking protocol clearly identified the endogenous circadian rhythm of HR. Although HR's circadian profile is phase locked with rectal temperature's, the authors point out that it is phase advanced by about 1 hour, which indicates it is not driven by CBT and can be considered an independent physiological parameter with evidence of endogenous circadian fluctuations.

A few years later, Krauchi *et al.* studied the link between distal vasodilation and sleep-onset latency, this time under constant-routine protocol modified to allow nocturnal sleep [22]. The conclusion of this study suggested that the distal-to-proximal skin temperature gradient was the best predictor for sleep initiation in 18 healthy adult males compared to HR, CBT and its rate of change, and even melatonin onset. They argue that despite HR and natural melatonin levels displaying circadian fluctuations, it is the circadian changes in thermoregulation which functionally relate to sleep propensity (i.e. the "pressure" of sleep). The takeaway for the purposes of this thesis is that HR was now also shown to display circadian fluctuations when nocturnal sleep is included in the constant routine protocol and not just when or if a subject stays up all day and all night, like could be argued from just their previous study.

In 2011, Boudreau *et al.*, which includes Dr. Boivin, investigated the circadian rhythm of HR and HRV [9]. They did this by monitoring RR intervals (via automatic detection software on 200Hz ECG), alongside CBT, salivary cortisol and urinary 6-sulfate-melatonin (UaMt6s; the main melatonin metabolite) as references for characteristic circadian markers. The study involved eight healthy young adults under a controlled protocol called the ultradian sleep-wake cycle (USW) procedure conducted over 72 hours. **Figure 5** below describes the USW procedure with an illustration showing alternating 60-minute wake episodes in dim light and nap opportunities in complete darkness. Throughout the procedure, subjects remained in bed in a semi-recumbent position and received isocaloric meals during each wake episode in order to limit the effect of active-rest and eating-fasting *zeitgebers*. RR interval values, standard RR interval low frequency

(LF: 0.04-0.15 Hz) and high frequency (HF: 0.15-0.40 Hz) bands relative power (separated via discrete wavelet transform), and the LF:HF ratio were the parameters considered in evaluating the correlation between HRV and circadian markers. The RR interval HF component is considered to reflect parasympathetic modulation of the cardiac system, whereas the LF:HF ratio is considered to reflect the sympathovagal balance [23]. The results of this study are presented in **Figure 6** and display significant circadian rhythms. Correlation coefficients and phase relationships with respect

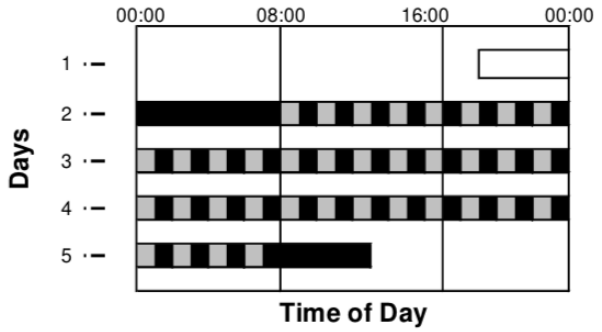

Figure 5: Ultradian sleep-wake cycle (USW) procedure. Following a baseline awake and then 8 hour sleep episode (days 1-2), subjects began an USW procedure for 72 hours (days-2-5). White bars represent waking hours in 150 lux, grey bars represent waking episodes in dim light (<10 lux) and black bars represent sleep episodes in total darkness (<0.3 lux). [9]

Table 1: Maximal correlation (units expressed in decimal hours) and correlation coefficient between the HRV parameters, and salivary cortisol and UaMt6s. A positive phase difference indicates an advance of the HRV parameter series compared to the hormonal series. ∗ indicates phase difference significantly different than zero ($p \leq 0.05$). Values are expressed as mean ± SE of the mean. [9]

	Salivary cortisol		*UaMt6s*	
	Phase Difference (h)	Correllation Coefficient	Phase Difference (h)	Correllation Coefficient
RR interval	6.25 ± 0.70 *	0.375	2.57 ± 0.84 *	0.483
LF power	5.75 ± 1.53 *	0.391	2.57 ± 1.36	0.474
HF power	8.25 ± 1.91 *	0.388	1.71 ± 2.11	0.478
LF:HF ratio	2.25 ± 2.09	0.439	-0.86 ± 1.62	0.428

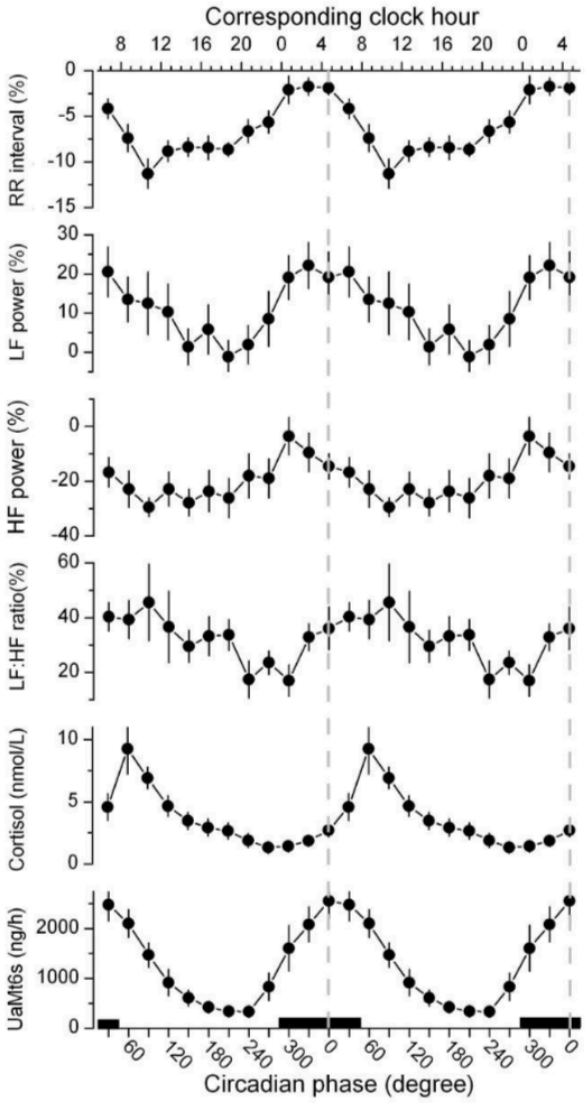

Figure 6: HRV and hormonal data during the USW procedure double-plotted over two days. HRV are expressed as percentage of the total 24-hour mean. Black bars at the bottom of graph represent time of projected habitual nocturnal sleep episodes. Values are expressed as mean ± SE of the mean. [9]

to cortisol and UaMt6s secretions, respectively, are recorded in **Table 1**. In summary, mean RR interval circadian profile is phase advanced with respect to salivary cortisol and UaMt6s rhythms and is higher at night compared to daytime (i.e. lower mean HR at night compared to day). The specific results of the LF, HF and LF:HF rhythms only conclusively suggest that the variation in parasympathetic modulation of the heart (HF power) is phase locked with UaMt6s secretion and phase advanced relative to salivary cortisol secretion, meaning maximum parasympathetic modulation habitually occurs at the onset of sleep.

In a latter study conducted by the same group under the same USW procedure, the researchers were able to show that circadian rhythm in HRV from fifteen healthy young adults does indeed correlate with increased cardiac sympathovagal response specifically at the time of awakening in the morning [16]. They did this by separately analyzing the HRV parameter series recorded during wake episodes from those recorded during nap episodes of the USW procedure. They concluded that interactions between sleep-wake dependency and circadian processes contribute towards the increased sympathovagal response to awakening in the morning such that it is shown to be phase locked with approximately zero lag to the morning cortisol peak. They go on to explain how this interaction may very well be the reason why adverse cardiac events, such as myocardial infarction, ventricular tachycardia, ventricular fibrillation and sudden cardiac death are predominantly reported in the morning hours [24][25].

2.3 Methods

2.3.1 Subject

One healthy adult male (25 years of age, 81 kg) is the sole subject of this study upon which all data was recorded on.

2.3.2 Monitoring Devices

The Polar M600 smartwatch, presented in **Figure 7**, was used as the primary device which derives wrist-based HR from an unknown proprietary algorithm converting wavelet recordings from the watch's photoplethysmography (PPG) sensors into units of beats per minute (bpm) at a rate of 1 Hz. The Hexoskin smartshirt, presented in **Figure 8**, was used as a secondary device in order to compare with and validate the smartwatch-derived HR. Similar to the Polar, the Hexoskin

also outputs HR recordings at a rate of 1 Hz, however derived from dry electrodes embedded in the garment producing an ambulatory ECG signal. **Table 2** summarizes and compares selected details about the two devices used in this study.

Table 2: Technical details of monitoring devices.

	Polar M600	Hexoskin
Type of wearable device	Smartwatch	Smartshirt
Sensors	6-LED PPG Accelerometer Gyroscope Ambient light sensor Vibration motor Microphone	Analog 2-lead 256Hz ECG Analog 3D 64Hz Accelerometer Analog dual-channel 128Hz breathing sensors
Raw sensor data available?	No	Yes
HR output	Derived from: Wrist-based PPG Output rate: 1 Hz Units: bpm	Derived from: 1-lead ECG Output rate: 1 Hz Units: bpm
Health/activity data provided	Activity tracker (steps, distance, cadence, speed/pace, calories burned, GPS). Sleep tracker (timing, amount quality). Wrist-based heart rate measurements (HR, HRmax, HR zones).	Activity tracker (steps, distance, cadence, speed/pace, peak acceleration, calories burned). Sleep tracker (timing, amount quality, position, resting HR and breathing rate). Cardiac measurements (QRS events, R-R interval, HR, HRmax, HR zones, HR recovery, resting HR, HRV). Respiratory measurements (breathing rate, tidal volume, minute ventilation, VO2max).
Battery life	24 to 36 hours	12 to 14 hours

Figure 7: Polar M600 smartwatch

Figure 8: Hexoskin smartshirt

2.3.3 Experimental Protocol

This study was split into two recording periods:

Recording Period 1- Data recorded over 15 days between Dec. 21, 2016 and Jan. 6, 2017.

Recording Period 2- Data recorded over 19 days between Feb. 19, 2017 and Mar. 10, 2017.

Data collection followed a strict protocol in order to (i) maximize data collection over as many hours of a day as possible, (ii) efficiently use full battery life of both devices, and (iii) to document as many significant activities/events to be used as reference when processing and analyzing the data. Below is a list of the tasks involved in the data acquisition protocol for this study:

1. Charge Polar to full battery.
2. Put Polar on and commence HR recording.
3. Charge Hexoskin to full battery.
4. Put Hexoskin on, plug in battery and begin recording.
5. Document (e.g. as a note file on smartphone) a journal log of all main activities/events that occur during the recording period in as much detail as possible. Below is a list of main events that needed to be recorded, however whatever else the subject felt should be documented was:
    5.1. Time of Waking up.

5.2. Time of going to Sleep.
(Considered to be when the subject is in bed, put all devices away (meaning phone, computer, iPad, etc.) and is ready to close their eyes and go to sleep).

5.3. Time + summary of morning/evening routine
(e.g. bathroom, shower, rushing to get out).

5.4. Time + cups/size of Coffee drink.

5.5. Start time + end time + portion + type of Food eaten
(e.g. 1 bowl of salad and 1 hamburger with spicy sauce).

5.6. Start time + end time + type and intensity of all Physical activity
(includes long walks, walking up/down stairs, gym workouts, sports, moments of heavy lifting).

5.7. Time + type of significant emotional stress
(e.g. rushing in the morning, school/work/family stress, excitement, anger, nervousness).

6. Recording from the Hexoskin would continue without interruption until either the battery died or the subject took a shower. When the subject became aware the battery died, they would right away charge it to full battery and continue recording once fully charged.
7. Recording from the Polar would continue without interruption until the battery died except for 15-20 min every morning after the subject woke up. When the subject became aware the battery died, they would right away charge it to full battery and continue recording once fully charged.
8. No research recordings took place from Friday sundown to Saturday nightfall.

Over Recording Period 1, both Polar and Hexoskin HR signals were recorded. Over Recording Period 2 only Polar HR signal was recorded.

2.3.4 Analysis

2.3.4.1 Polar HR validation with Hexoskin HR

After aligning the start time of all Polar and Hexoskin HR data files over the recording periods, on segments where both Polar and Hexoskin HR data were available the difference and cross-correlation between the two signals were investigated. The premise is such that Polar's wrist-

based PPG-derived HR signal needs to be validated with Hexoskin's chest-based ECG-derived HR signal prior to developing subject-specific models for circadian fluctuations of HR and further circadian analyses derived solely from a smartwatch with PPG-derived HR recording available.

To begin with, the cross-correlation between the two HR signals over segments where both were collected was estimated and the delay at the maximum correlation was used to align the two signals for further analysis. From there, for each segment, the (1) error between the two HR signals was computed by subtracting the Polar HR from Hexoskin HR (error signal also referred to as the residual), (2 and 3) the probability distributions of each signal and of the residual were computed, and (4 and 5) power spectrums for each signal and the residual signal were computed. Together, these results are used to analyze and understand the temporal and frequency domain comparability or lack thereof between both devices' HR outputs.

(i) Mean percent error (MPE), (ii) root-mean squared-error (RMSE), and (iii) maximum value of cross-correlation coefficient function were summarized for each segment of both Polar and Hexoskin HR data and then averaged over all segments to quantify the comparability of Polar's HR signal with Hexoskin's. MPE and RMSE calculations are presented in **Equations 1 and 2**, respectively:

$$MPE = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{h_i[n] - p_i[n]}{h_i[n]}\right| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(h_i[n] - p_i[n])^2} \tag{2}$$

where h_i is the Hexoskin HR signal sampled at 1 Hz and p_i is the Polar HR signal also sampled at 1 Hz, which are both collected over the same segment i of duration equal to N samples long. MPE and $RMSE$ were calculated for each segment individually.

2.3.4.2 Subject-specific model for characterizing circadian fluctuations of HR

A model will be proposed to characterize free-living environment circadian fluctuations of HR derived from the Polar smartwatch over the full recording period (Recording Periods 1 and 2).

First, all HR data was segmented into their respective calendrical date's 24-hour timeline and then the mean of all HR data points available at each clock second was computed in order to output a mean HR signal over the 24-hours in a day. The mean HR signal was computed from all available HR data points, however was then recomputed after exogenously elevated or depressed HR segments due to external factors were cut out. The external factors considered to non-endogenously elevate or depress HR were namely exercise and all-nighters where the subject purposefully stayed up. The start and end time of these activities were determined from the detailed journal log the subject kept. HR intervals during sleep was not removed as an external factor since sleep is a natural process of healthy daily life and was decided to be included in this study's analysis as an inevitable external factor that will present itself over 24-hour cycles. Contrarily, exercise is not an event that necessarily occurs over every 24-hour cycle in a normal life (however, may be an activity which people strive to do every day or so), therefore was considered as extenuating circumstances which can contribute to a shift in the subject's circadian rhythm. If the subject exercised daily or even every second day at a consistent time, this may have been reconsidered, however, as the journal log and data show, this subject exercises once or twice a week at most. HR data over such external factors was manually removed by visually detecting when the HR elevation/depression began and ended with respect to baseline.

The mean HR signal after intervals of external factors were removed, also to be referred to as the circadian HR profile, was then smoothed with a low pass filter (LPF) at different cut-off frequencies until an acceptable cut-off was identified. Along with the mean HR signal, the SE at each time point was also computed by dividing the standard deviation of HR values at the specific time point by the square root of the number of HR values available at the respective time point. The LPF with same cut-off frequency was also applied to the SE signal and was used as a ± boundary of the circadian HR profile for analysis purposes.

Once the smoothed circadian HR profile was computed, different parameterization methods to characterize the profile were proposed. The four parameterization models considered and their mathematical equations are listed below:

1) Polynomial:
$$f(x) = a_0 + \sum_{n=1}^{N} a_n x \tag{3}$$

2) Sum of Sines: $$f(x) = \sum_{n=1}^{N} a_n \sin(b_n x + c_n) \quad (4)$$

3) Fourier Series: $$f(x) = a_0 + \sum_{n=1}^{N} a_n \cos(nxw) + b_n \sin(nxw) \quad (5)$$

4) Gaussian: $$f(x) = \sum_{n=1}^{N} a_n \exp\left(-\left(\frac{x - b_n}{c_n}\right)^2\right) \quad (6)$$

The decision criteria for selecting the best model are:

i. Lowest number of coefficients
ii. Highest adjusted R^2 value
iii. Lowest RMSE

The model with the best combination of statistics satisfying the decision criteria will be selected and used to define this study's subject-specific model for characterizing circadian fluctuations of HR.

2.4 Results

Figure 9 presents all Polar HR signals collected over Recording Periods 1 and 2 concatenated with respect to date and time and separated into weeks from Sunday to Saturday; Weeks 1 and 2 comprise Recording Period 1 and Weeks 3 to 5 comprise Recording Period 2. Similarly, **Figure 10** presents all Hexoskin HR signals, which were only collected over Research Period 1. The error between Polar and Hexoskin HR signals were computed on all segments with both signals available and the residual plotted in **Figure 11**. Significant external factors, such as sleep and exercise intervals, are identified in these plots by bar lines over the duration they occurred at as per journal log entries.

Appendix A presents figures of all the raw Polar HR data segmented into calendrical days for Recording Periods 1 and 2, respectively.

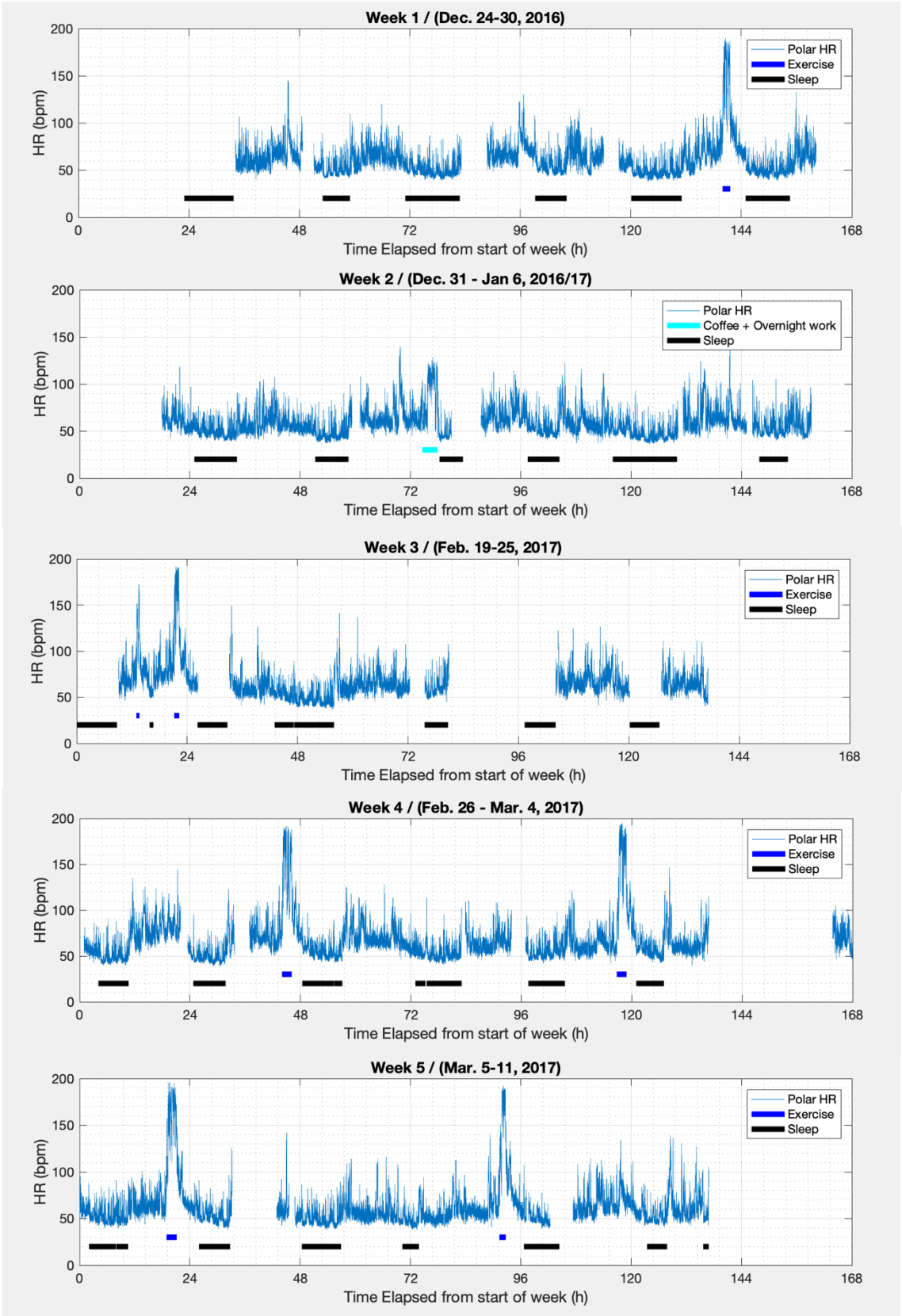

Figure 9: Polar HR signals concatenated and separated into weeks (Sunday to Saturday) over Recording Period 1 (Weeks 1 and 2) and Recording Period 2 (Weeks 3-5).

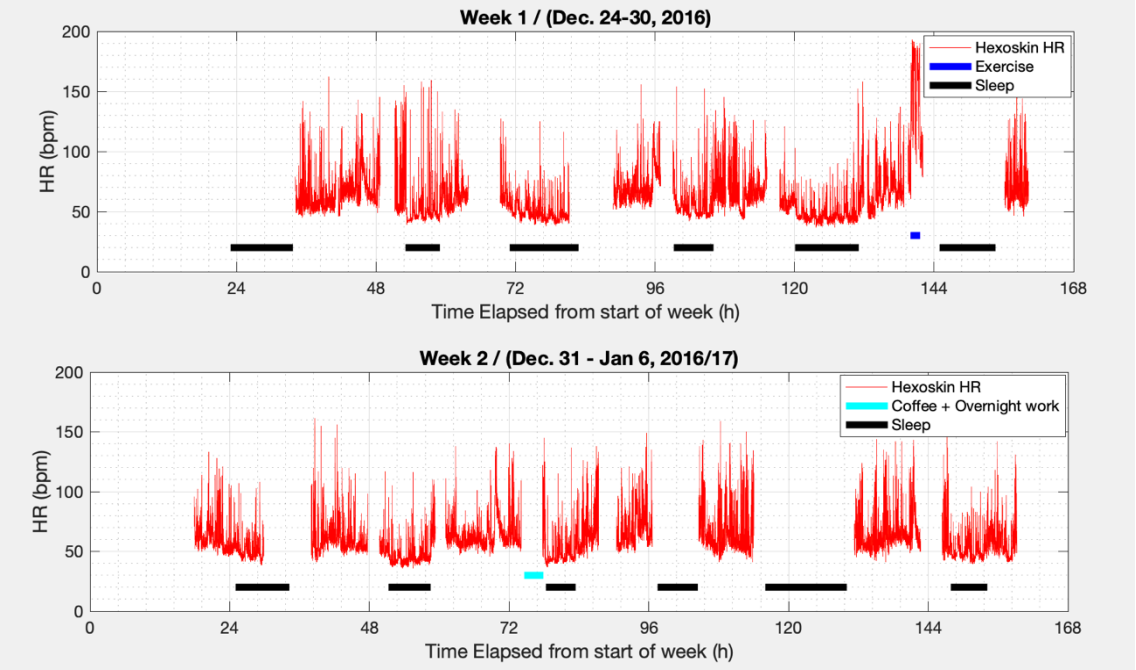

Figure 10: Hexoskin HR signals concatenated and separated into weeks (Sunday to Saturday) over Recording Period 1 (Weeks 1 and 2).

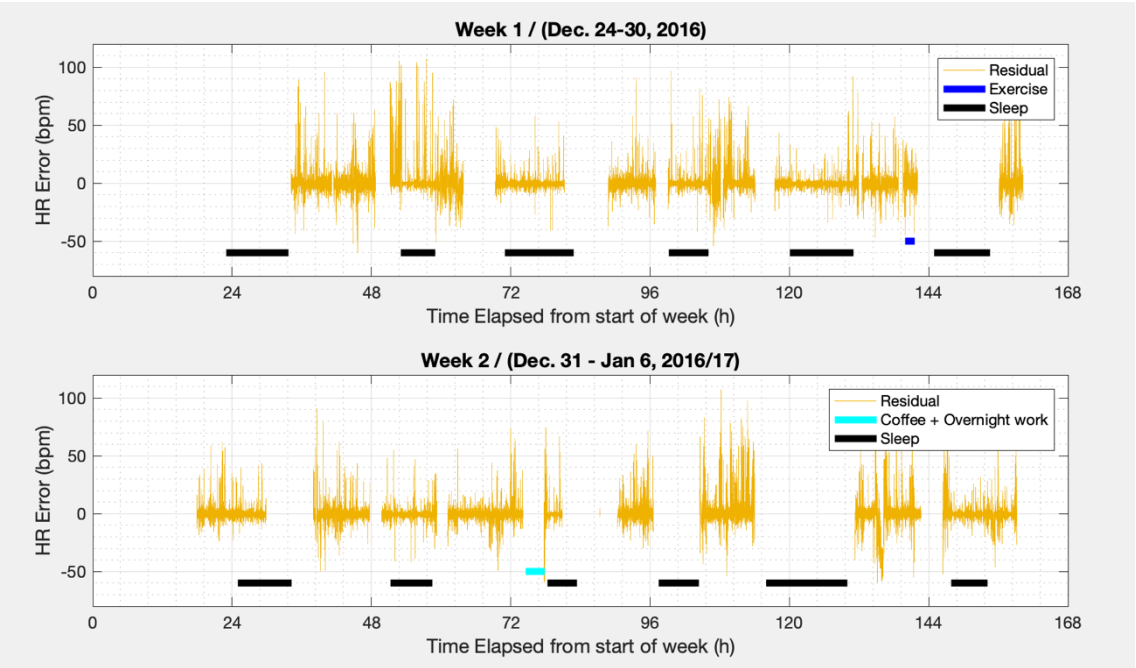

Figure 11: HR Error between Polar and Hexoskin concatenated and separated into weeks (Sunday to Saturday) over Recording Period 1 (Weeks 1 and 2).

In **Table 3**, summary statistics MPE, RMSE and cross-correlation on each segment and the overall average (and standard deviation) between Polar's HR and Hexoskin's are presented. 25

segments of at least 2 hours in duration were available for analysis. **Figure 12** and **Figure 13** demonstrate the results generated to extract the summary statistics presented in **Table 3** and further analysis plots comparing the two HR signals on two selected segments, respectively. For illustration purposes, the former is one of the segments yielding highest combined summary statistics and the latter yielding one of the lowest.

Table 3: Summary Comparison of Polar HR signal with Hexoskin HR signal over 25 corresponding segments during Recording Period 1.

	Duration [h]	MPE [%]	RMSE [bpm]	Cross-Corr.
Segment 1	13.00	8.26	12.13	0.65
Segment 2	11.71	9.17	11.27	0.73
Segment 3	6.97	6.46	9.28	0.63
Segment 4	7.20	6.23	8.38	0.80
Segment 5	12.65	9.22	14.07	0.49
Segment 6	5.53	2.43	2.43	0.93
Segment 7	4.59	2.48	3.75	0.75
Segment 8	8.16	4.01	5.47	0.91
Segment 9	7.44	3.22	6.27	0.83
Segment 10	5.52	6.67	10.00	0.64
Segment 11	14.45	2.83	4.31	0.89
Segment 12	6.30	6.00	7.45	0.83
Segment 13	2.54	4.21	7.89	0.98
Segment 14	4.07	9.98	12.55	0.49
Segment 15	10.48	3.48	4.98	0.88
Segment 16	9.16	5.53	7.26	0.78
Segment 17	2.26	2.97	3.36	0.90
Segment 18	7.23	3.48	4.20	0.90
Segment 19	13.00	4.25	6.25	0.88
Segment 20	3.20	6.08	10.64	0.74
Segment 21	6.14	5.07	7.42	0.83
Segment 22	9.13	9.25	13.05	0.59
Segment 23	11.37	10.91	13.77	0.56
Segment 24	8.79	4.29	6.29	0.80
Segment 25	3.94	5.57	7.94	0.81
Avg. (Std.)	**7.79** (3.52)	**5.68** (2.53)	**8.02** (3.42)	**0.77** (0.14)

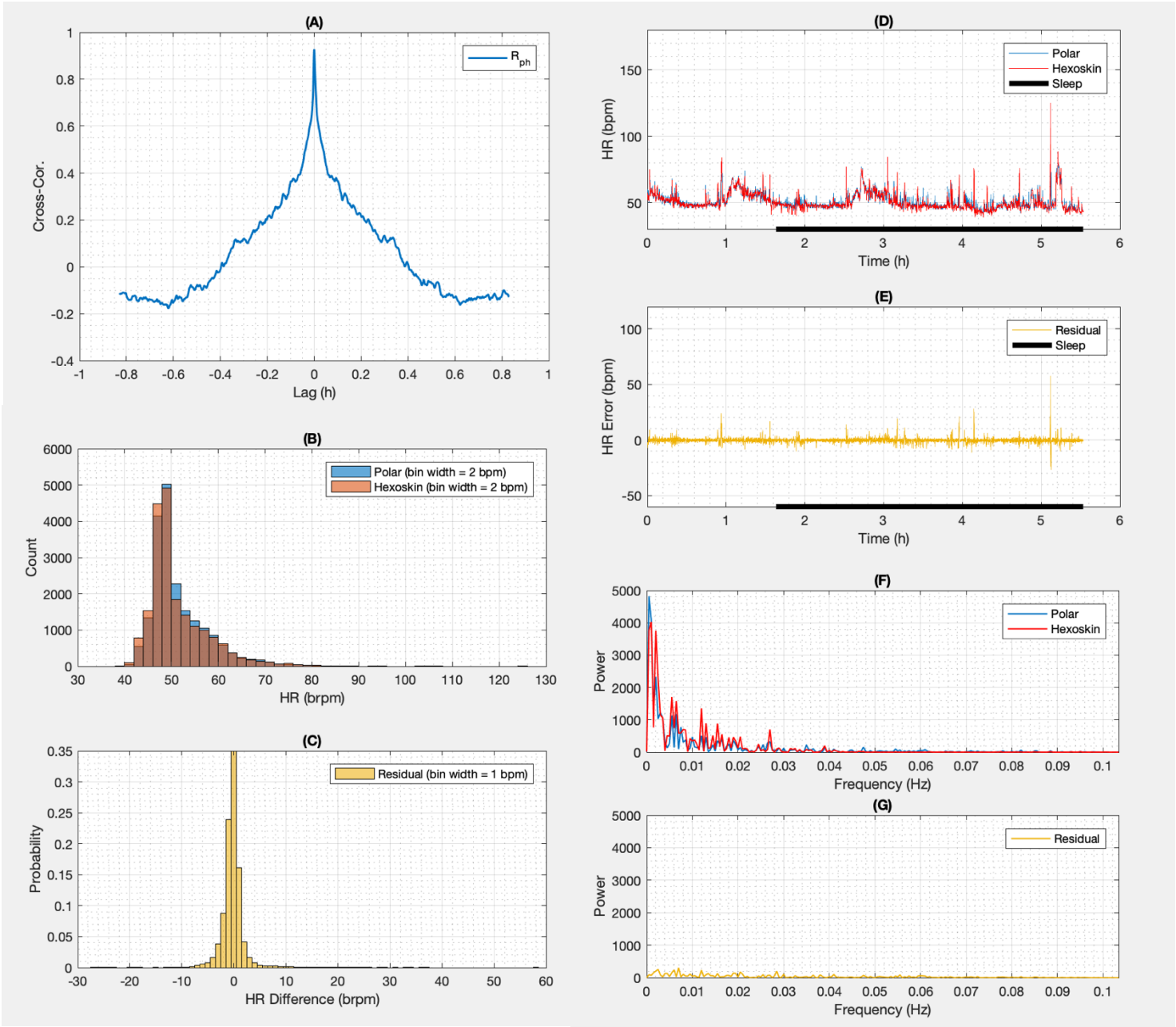

Figure 12: Polar/Hexoskin HR signal comparison over Segment 6 corresponding to the recording on Dec. 26-27, 2016. **(A)** Cross-correlation coefficient function (R_{ph}) of Polar HR signal (p) with respect to Hexoskin HR signal (h) after max. correlation aligned to zero lag; **(B)** Frequency distribution of Polar HR values and Hexoskin HR values, respectively, with bin sizes of 2 bpm; **(C)** Probability distribution of the residual between Hexoskin HR and Polar HR with bin sizes of 1 bpm; **(D)** Polar and Hexoskin HR time-series, respectively, and the duration of which was during evening sleep (black bar); **(E)** Residual (Hexoskin - Polar) time-series and sleep interval; **(F)** Power spectrum of Polar and Hexoskin HR signal, respectively, zoomed-in to range between 0 Hz and approx. 0.1 Hz; **(G)** Power spectrum of residual zoomed-in to range between 0 Hz and approx. 0.1 Hz.

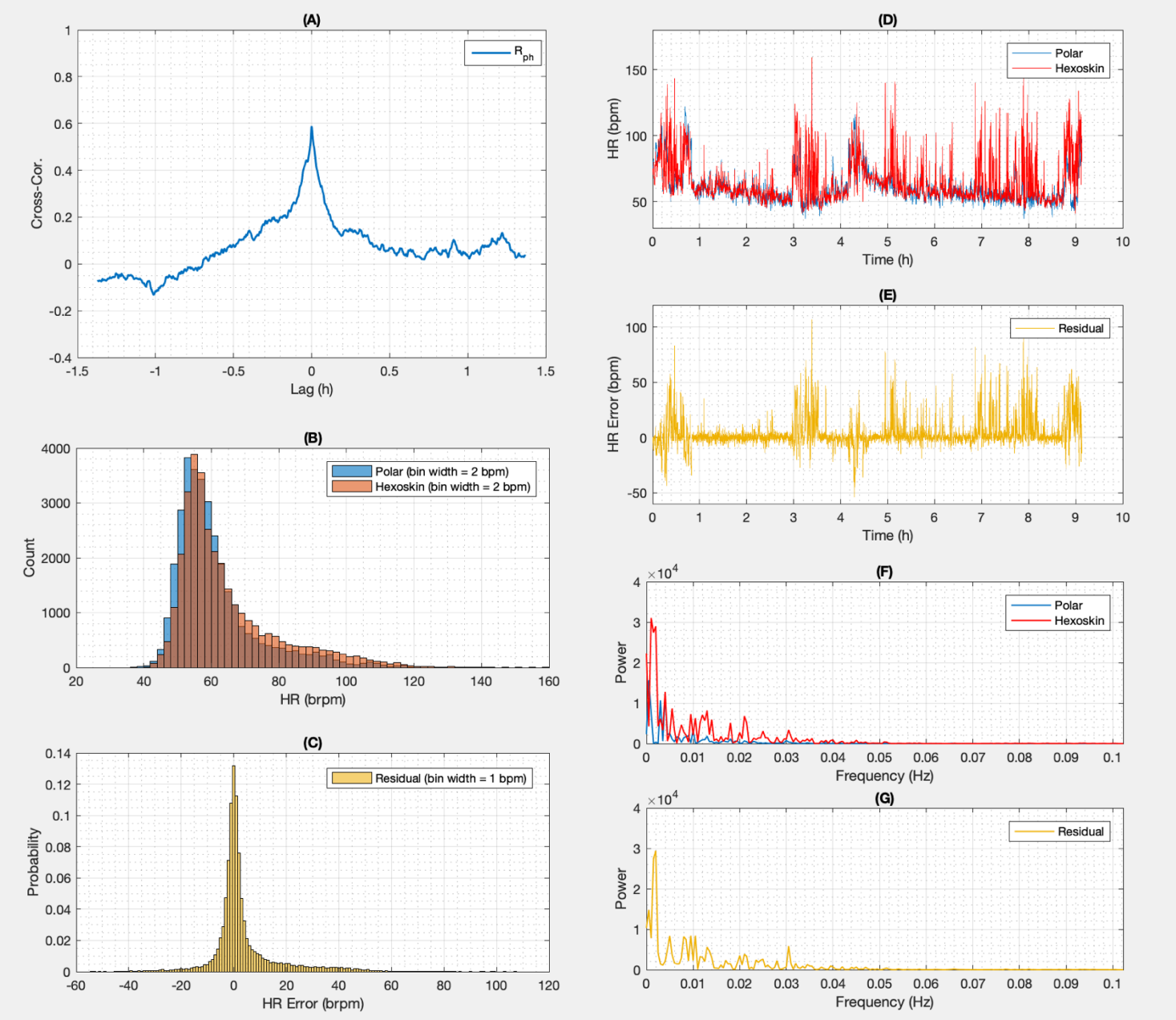

Figure 13: Polar/Hexoskin HR signal comparison over Segment 22 corresponding to the recording on Jan. 4, 2017. **(A)** Cross-correlation coefficient function (R_{ph}) of Polar HR signal (p) with respect to Hexoskin HR signal (h) after aligned for max. correlation at zero lag; **(B)** Frequency distribution of Polar HR values and Hexoskin HR values, respectively, with bin sizes of 2 bpm; **(C)** Probability distribution of the residual between Hexoskin HR and Polar HR with bin sizes of 1 bpm; **(D)** Polar and Hexoskin HR time-series, respectively, all of which was during awake hours; **(E)** Residual (Hexoskin - Polar) time-series; **(F)** Power spectrum of Polar and Hexoskin HR signal, respectively, zoomed-in to range between 0 Hz and approx. 0.1 Hz; **(G)** Power spectrum of residual zoomed-in to range between 0 Hz and approx. 0.1 Hz.

Figure 14 presents the results of this subject's mean HR (Polar) calculated as the average HR at each second (since 1 Hz HR outputs) over the 24 hours of a day over Recording Periods 1 and 2, which came out to 34 days of HR data. Although most days did not have HR data at every second, the ensemble of HR data did have more than just one at each second of a 24-hour clock. **Figure 14.A** considers averaging of all HR data whereas **Figure 14.B** removed segments from the averaging where external factors, except for sleep, contributed towards exogenously elevated or depressed HR values, such as exercise as discussed earlier. Results and analysis moving forward considers the mean 24-hour HR signal with such exogenously elevated/depressed HR values removed before averaging (i.e. **Figure 14.B**) – the circadian HR profile.

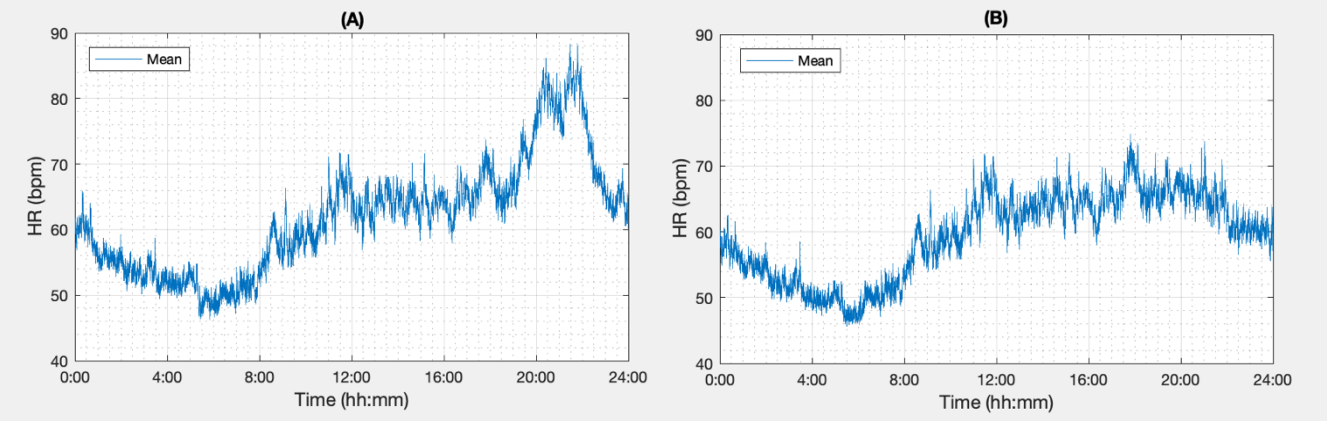

Figure 14: Mean across all Polar HR values recorded at the same time of day (hh:mm) over Recording Periods 1 and 2 (34 days total), referred to as the subject's circadian HR profile. **(A)** All HR data included; **(B)** Significant external factors, except sleep, contributing to exogenously elevated or depressed HR values (e.g. exercise) manually removed from averaging.

Figure 15 illustrates Low Pass cutoff frequencies $1x10^{-4}$ Hz, $1x10^{-5}$ Hz and $5x10^{-5}$ Hz for smoothing the circadian HR profile. When comparing the three, at the lower end $1x10^{-4}$ Hz cutoff frequency (**Figure 15.A**) the smoothed signal appears to still capture higher frequency oscillations that most likely represent variation as a result of averaging as oppose to a general 24-hour endogenous HR fluctuation profile; at the higher end $1x10^{-5}$ Hz cutoff frequency (**Figure 15.B**) the smoothed signal evidently misses a lot of the higher frequency components and thus misses the local maximums and minimums present in the signal and which contributes to the characteristic circadian HR profile as brought down in literature. However, at cutoff frequency of $5x10^{-5}$ Hz (**Figure 15.C**), a good balance between the two extremes is evident, which becomes further

apparent when considering where the mean signal lies between the ± 1x and 2x SE boundaries compared to the other two cutoff frequencies. As a result, the $5x10^{-5}$ Hz low pass cutoff frequency is selected to smooth the circadian HR profile for further analysis in proposing a subject-specific model that characterizes circadian fluctuations of HR.

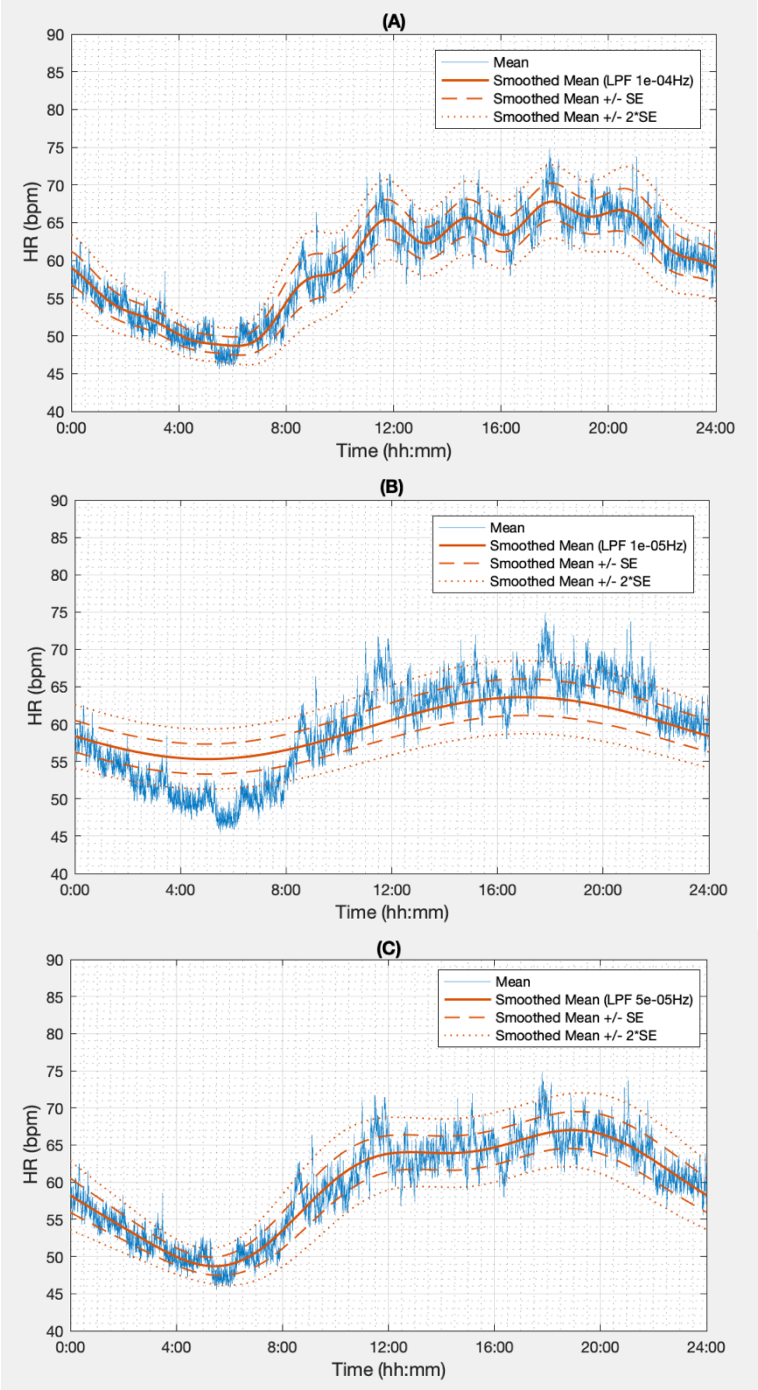

Figure 15: Smoothing of Mean 24-hour HR signal with different low pass cutoff frequencies: **(A)** $1x10^{-4}$ Hz; **(B)** $1x10^{-5}$ Hz; and **(C)** $5x10^{-5}$ Hz, which is the selected LPF.

Once the smoothing operation was selected, key anchors (1) minimum, (2) interval of maximum increase and (3 and 4) two local maximums were identified. This was achieved by computing the derivative of the circadian HR profile as illustrated in **Figure 16.A** and then identified on the actual circadian HR profile in **Figure 16.C**. The probability distribution of when the subject was sleeping over Recording Periods 1 and 2 was computed and plotted in **Figure 16.B**. From this, a line bar separating probability distributions into six categories is plotted on a grey scale at the bottom of **Figure 16.B** was included as well at the bottom of **Figure 16.C**. This sleep

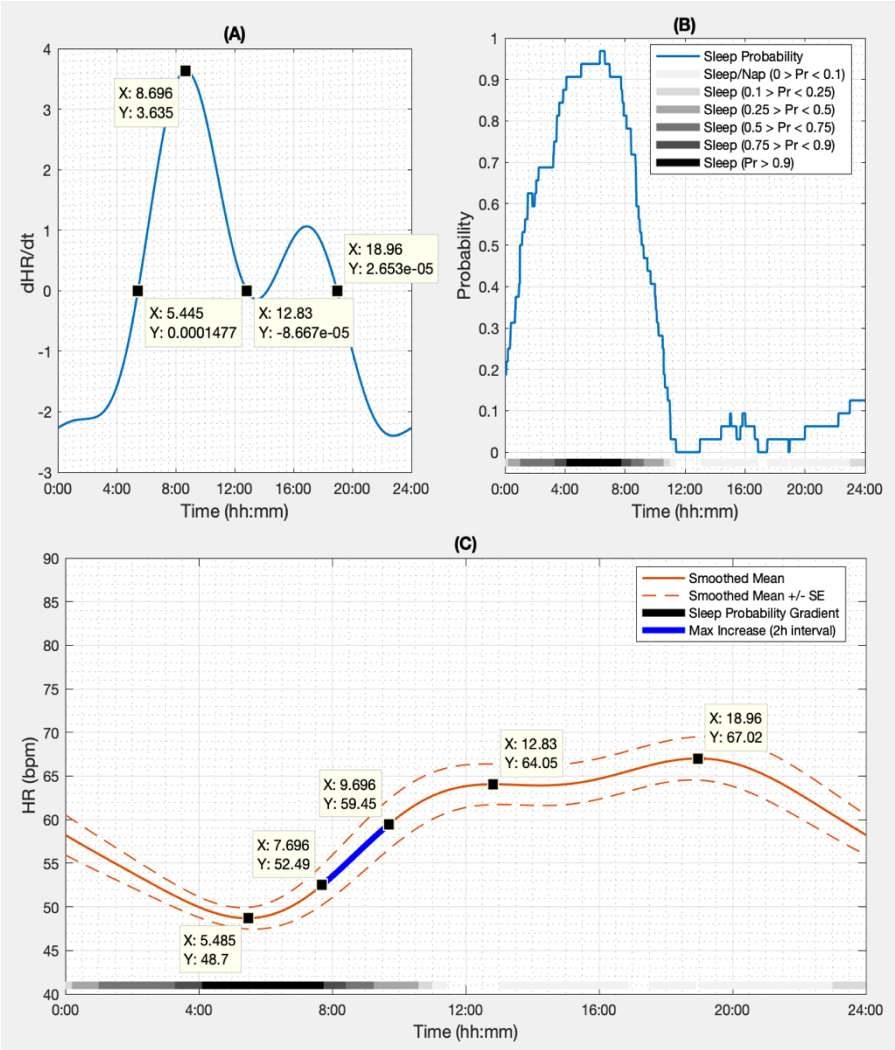

Figure 16: Anchors derived and sleep distribution defined in order to provide context to the circadian HR profile. **(A)** Derivative of Smoothed Mean signal ($5x10^{-5}$ Hz LPF) with datatips at HR circadian anchors; **(B)** Probability distribution of subject's sleeping intervals, including naps with grey scale gradient line bar of the sleep distribution; **(C)** Smoothed Mean and 1x SE boundaries of the circadian HR profile with gradient line bar of sleep distribution and key anchors marked by datatips (from left to right): Minimum, Interval of Maximum Increase, Local Maximum 1 and Local Maximum 2.

probability gradient provides valuable context to when these anchors occur with respect to the subject's sleep-wake schedule over the full recording period from which the circadian HR profile was derived from. Values and time of day for the HR circadian anchors are presented in **Table 4** below.

Table 4: Key anchors identified in the circadian HR profile. Values expressed as smoothed 24-hour mean ± SE.

	Definition	Time [hh:mm]	Value [bpm]
Anchor 1	Minimum	05:29	48.7 ± 0.2
Anchor 2	Interval of Maximum Increase	07:42-09:42	+7.0 ± 2.9
Anchors 3 and 4	Local Maximum 1	12:50	64.1 ± 2.3
	Local Maximum 2	18:58	67.0 ± 2.5

The parameterization methods considered to mathematically model the circadian HR profile and their model order, adjusted R^2, RMSE and number of coefficients are presented in **Table 5**. Based on the decision criteria of (i) lowest number of coefficients, (ii) highest adjusted R^2, and (iii) lowest RMSE, the 3rd order Fourier Series parameterization method is selected as the best model to characterize this subject's circadian fluctuations of HR. Although the 3rd and 4th order Sum of Sines methods also yield good fit statistics, the 3rd order Fourier Series yields 1.0000 adjusted R^2 with less number of coefficients and with comparable RMSE, therefore Fourier Series model was selected.

Table 5: Parameterization methods and associated statistics used in selecting the best model for characterizing subject-specific circadian fluctuations of HR.

Parameterization Method	Model Order	Adjusted R^2	RMSE	No. of Coefficients
Polynomial	8	0.9953	4.208×10^{-1}	9
	9	0.9991	1.878×10^{-1}	10
Sum of Sines	2	0.9143	1.791×10^{-1}	6
	3	0.9908	5.867×10^{-1}	9
	4	1.0000	2.575×10^{-2}	12

	2	0.9895	6.263×10^{-1}	6
Fourier Series	**3**	**1.0000**	**3.486×10^{-2}**	**8**
	4	1.0000	4.705×10^{-14}	12
	2	0.9696	1.066×10^{0}	6
Gaussian	3	0.9995	1.426×10^{-1}	9
	4	0.9963	3.742×10^{-1}	12

The equation of the selected 3rd order Fourier Series model with its coefficient are presented in **Equation 7** below, where $\widetilde{HR}$ is the model estimated circadian HR profile in units of bpm, and t is the time in units of hours which loops in the range of [0,24[.

$$\widetilde{HR}(t) = a_0 + \sum_{n=1}^{N=3} a_n \cos(ntw) + b_n \sin(ntw) \tag{7}$$

Coefficients:

$$a_0 = 59.45 \qquad b_1 = -7.986$$
$$a_1 = -2.187 \qquad b_2 = -1.556$$
$$a_2 = 1.614 \qquad b_3 = 0.8772$$
$$a_3 = -0.6393$$

$$w = 0.2614$$

2.5 Discussion

Before evaluating endogenous fluctuations of HR derived from the Polar M600 smartwatch, the reliability of its HR over all hours of the day needed to be established. This was done by monitoring HR from another wearable device, the Hexoskin smartshirt, which moreover derives HR from a different signal; ambulatory ECG as oppose to PPG. The main limitations in regard to comparing both HR outputs over multiple consecutive days was (i) the different battery

life of each device, and (ii) data alignment. The former was able to be lessened by disconnecting Bluetooth connectivity on the Polar, increasing its uninterrupted HR monitoring from around 24 hours to just about 36 hours and by optimizing recharging time for both devices to exactly the amount of time needed to bring the device close to 100%. Hexoskin battery lifetime was found to be nearly half that of Polar; however, because Polar Bluetooth was disconnected and recharging time was optimized, majority of the time Hexoskin HR was recorded so to was Polar's. This can be seen when comparing the HR signals over Weeks 1 and 2 from both devices (**Figure 9** and **Figure 10**) such that only a few short intervals in each week Polar HR is recorded and Hexoskin is not. Battery life for wearable devices is still a concern today when the aim is to record multiple, uninterrupted, consecutive days as would be needed for circadian rhythm studies. However, with each year the field advances, so too does the battery lives. As for the second limitation, since both device HR outputs are sampled at 1 Hz exactly, this made things simpler. For analysis purposes, data alignment was mitigated by first computing the cross-correlation coefficient of each segment and then adjusting the timestamp of the Hexoskin HR outputs with what was the point of maximum correlation's lag. It was observed that Hexoskin's timestamp was anywhere between 4 to 7 seconds advanced compared to Polar's HR signal. This can be explained in part due to the physiological delay between HR derived from the QRS event in ECG signal compared to HR, rather pulse rate, derived from PPG waveforms recorded at the wrist. Studies have shown that this delay, referred to the pulse transit time (PTT), can range anywhere between 180 and 330 ms when detected at the wrist/finger [26],[27]. Therefore, the delay in Polar's HR signal can surely be assumed to be more significantly associated with Polar's proprietary algorithms (unknown) used to process the PPG signal and technical or processing delays associated with sending and timestamping outputs.

Once Hexoskin and Polar data was aligned, segments with both data available were collected and the residual computed: Hexoskin HR minus Polar HR at each time point (**Figure 11**). From these results over Weeks 1 and 2, it can be seen that the residual predominantly hovers around 0 bpm, with intermittent and relatively sharp errors every so often especially when the subject is not sleeping. Upon further analysis, it can be seen in **Figure 11** that the sharp intermittent errors weigh heavier on the positive y-axis than on the negative, meaning most of the time there are significant differences between the two HR signals, Hexoskin's is significantly higher. This can be seen as well in **Figure 9** and **Figure 10** with close attention. **Table 3** summarizes the

comparison statistics computed on all segments with both HR signals available (of at least 2 hours) and yields an average MPE of 5.68%, average RMSE of 8.02 bpm, and average correlation coefficient of 0.77 across all segments. Together, these results suggest high degree of comparability between both devices' HR output, and allow me to conclude for purposes of this study that Polar's HR signal is reliable, especially after the error between the two can be explained to be on the part of Hexoskin's signal and not Polar's.

To provide context as to why significant power at high frequency was observed form this study in Hexoskin HR signals compared to Polar HR signals, a little background on ECG, and ambulatory ECG signal from the Hexoskin specifically, needs to be discussed here. Hexoskin's HR is derived from electrodes embedded in the fabric of the smartshirt and produce an ambulatory ECG signals. It is well documented that ECG signals are challenged with several forms of artifact – physiological, experimental, and environmental – however, when it comes to ambulatory ECG signals, such as the one monitored by the Hexoskin, the main artifact corrupting the signal is due to electrode motion artifact [28][29]. Electrode motion artifact in the ECG signal produces R-peak-like artifacts corrupting any R-peak detection algorithm used to calculate HR and thus introduces significant error in HR output above certain signal-to-noise ratios (SNR). On one hand, either a *stricter* R-peak detection algorithm is implemented which results in relatively higher false negative (FN) rates and lower false positive (FP) rates, thus outputting erroneously low HR values. On the other hand, a *looser* R-peak detection algorithm can be implemented to achieve low FN rates however at the expense of relatively high FP rates, thus outputting erroneously high HR values. Hexoskin, like many other ECG monitoring devices, lean more towards the *looser* R-peak detection algorithm in order not to miss any true R-peaks and add other sections to the algorithm to mitigate FPs without compromising the true positive (TP) detection rates (e.g. recursive filtering [30]). It is therefore with reason to assume most of the high frequency, elevated HR values outputted in Hexoskin HR signals, which are significantly more prominent during wakeful hours compared to sleeping hours, is a consequence of deriving HR from ambulatory ECG and the common FP beat detections associated with it due to motion artifacts. To further evaluate this with respect to our study and the specific relationship observed between Hexoskin HR, Polar HR and residual HR, detailed analysis on Polar/Hexoskin segments were performed and results on two segments presented in **Figure 12** and **Figure 13**, respectively. The first presents results over segment 6 where the subject was preparing for nocturnal sleep and then sleeping (i.e. no-to-low

activity), and the second over segment 22 which consisted of regular awake hours of the day with intervals of low-to-moderate activity. Comparing panels **A** of these two figures, the no-to-low activity segment yields significantly higher correlation between the two signals than the low-to-moderate activity segment (0.93 compared to 0.59). This is evident when considering panels **D** and **E** with both HR plotted over each other and the residual plotted, respectively; segment 6 Hexoskin and Polar HRs closely follow each other and yield a roughly zero error throughout, whereas segment 22 Hexoskin HR continuously overshoots the Polar HR with high frequency and large positive spikes also apparent in the residual signal. Panels **F** further verifies this such that segment 22 Hexoskin HR power is more than double that of Polar's and has significant power in higher frequencies (between 0.01 and 0.03 Hz) not only compared to Polar's over the same segment but also compared to Hexoskin's over segment 6. In panels **G**, segment 6 residual appears as white noise (until about 0.1Hz where all power saturates to zero), whereas there is still significant power in segment 22 residual. Panels **B** and **C** present the overlaid probability count of each signal separated into 2 bpm bins and the probability distribution of the residual separated into 1 bpm bins, respectively. The Hexoskin probability count over segment 6 closely relates to the Polar's despite still having a few counts at higher HR values when Polar does not, nonetheless the residual probability distribution appears to resemble that of a normal or sharp student-t distribution function in the ± 10 bpm range. As for segment 22, the probability count of Hexoskin overshoots Polar's at higher HR values and surprisingly undershoots Polar's at lower HR values. Consequently, a nonsymmetric residual distribution with larger error at higher HR values is the result. Together, the analysis on these two segments as example confirms our initial observation that, in general, the Hexoskin outputs higher HR signal with larger high frequency components, especially during wakeful hours of the day, compared to Polar HR signal. With the understanding that Hexoskin's ambulatory ECG-derived HR signal inherently is susceptible to motion artifacts resulting in more FP beat detections and thus erroneously high HR values compared to Polar's HR signal at times, the comparability of the two signals outweigh the minor differences observed in the detailed signal analysis. Furthermore, it was the trend over a 24-hour scale which was important to validate for the Polar smartwatch's HR signal since circadian fluctuations is the focus of this research. Therefore, this study concludes that there is not enough evidence to suggest Polar's HR is unreliable and will be used with confidence when analyzing subject-specific circadian fluctuations of HR.

Once the smartwatch's HR signal was established as reliable, we can rely on earlier studies identifying endogenous, diurnal fluctuations of HR that it does exist and is identifiable under constant routine protocols with or without sleep or under alternating sleep-wake condition protocols (e.g. USW procedure) [9][11][16][22]. The main objective this study was thus to test whether circadian fluctuations of HR can be verified under complete free-living conditions or not from the primary wearable device. The results of this study argue that indeed it can be. Despite high frequency perturbations and gaps of data missing, visual inspection of **Figure 9**, which segments all Polar recordings into Sunday-Saturday weeks, clearly reveals a significant low frequency component with a period of roughly 24 hours. Therefore, what was done to create a 24-hour profile of this subject's HR was to take an average of all HR data available at each second of the 24-hour clock; **Figure 14.A** displays these results. Noticeably, between around 19:00 h to 23:00 h there is a large spike in mean 24-hour HR. This is because, as inferred from the subject's journal log, 5 days over the recording period the subject exercised in the evening around those times (**Figure 9** confirms this with blue bar lines indicating the interval of exercise). Specifically, the subject played basketball for 1 to 2 hours where his HR elevated to just under 190 bpm peak and stayed around the range of 150 to 180 bpm throughout the exercise. Once exercise finished, baseline endogenous HR resumed after up to even 4 hours of recovery. Depending on the intensity and duration of exercise, it is expected that HR recovery back to baseline varies and can take several hours after high intensity exercises because of the central response and local peripheral response components of the cardiovascular system [31]. Although exercise was not prohibited since the aim of this study was to monitor circadian rhythms under uncontrolled, natural conditions with no restrictions on activities, exercise intervals and other intervals confirmed from journal log that exogenously elevated or depressed HR, such as overnight caffeinated work, were excised from the data to create a new mean 24-hour HR profile. **Figure 14.B** illustrates the results which is argued to reflect only naturally occurring, diurnal, endogenous fluctuations. For this reason, sleep intervals was not removed since it naturally occurs on a daily basis for all healthy human beings unless purposely avoided (like the subject of this study did one night). Exercise though is not a prerequisite for daily life to the extent sleep is; regular exercise for multiple times a week is recommended in a healthy lifestyle, however there is no propensity to exercise like there is to sleep unless the individual accustomed themselves to exercise daily for many years and even then can relatively be easily stopped for prolonged periods without serious repercussions, as would be the

case for not sleeping (e.g. hallucination, motor function impairment, etc.). It is here where this study differentiates between two of the main *zeitgebers* after sunlight exposure when it comes to evaluating circadian fluctuations in free-living conditions: daily sleep-wake schedule is a human requirement; exercise is not per say and therefore treated as a significant external factor masking the endogenous HR fluctuations interested in analysing.

After the endogenous mean 24-hour HR signal was collected as defined, LPF at different cut-off frequencies were tested in order to best smooth the data while still preserving key anchors identified in the data and which have been brought down in literature. **Figure 15** summarizes this evaluation showing a relatively low cut-off frequency of $1x10^{-4}$ Hz in panel **A**, a relatively high cut-off frequency of $1x10^{-5}$ Hz in panel **B**, and one in the middle of $5x10^{-5}$ Hz in panel **C**, which was selected as the acceptable LPF cut-off frequency for further analysis of the circadian HR profile.

Value and time of day for the minimum and maximum HR values are two key anchors commonly set out to define when characterizing physiological signals on a circadian map [9][22][32]. CBT recorded rectally is a well-established parameter monitored alongside other parameters of interest in circadian studies since its nadir (i.e. minimum value) repeatably occurs in the early hours of the morning or just past the halfway mark of nocturnal sleep [33]. Therefore, anchors are also commonly expressed in units of time after (+) or before (-) the CBT nadir or in units of degrees after segmenting one days CBT nadir with the next day's CBT nadir into 360 degrees. This study only recorded HR, therefore times of anchors can only be expressed by their value and time of occurrence along our 24 hour clock. Besides for minimum and maximum HR value anchors, researchers sometimes also define the interval of maximum increase in HR since it strongly correlates with the time majority of adverse cardiac events have been reported to occur at [16]. The circadian HR anchors identified from this study are presented in **Table 4**, which were derived from the results presented in **Figure 16**. To further provide context of the circadian HR profile and before being able to fully discuss the four anchors identified, it was essential that sleep-wake schedule information was captured since it was not a controlled *zeitgeber* like the majority of studies published. To do this, a probability distribution of which hours the subject was sleeping over the full recording period was computed and results presented in **Figure 16.B** and then again

in **Figure 16.C** as a grey-scaled gradient bar line under the circadian HR profile. This is a novel representation of a subject's sleep-wake schedule when studying circadian rhythms in free-living conditions and proved to be valuable in our understanding of the proposed circadian HR profile, its SE boundaries and the anchors. HR circadian anchors were derived from taking the derivative of the smoothed mean 24-hour HR signal, as presented in **Figure 16.A**. Minimum HR of 48.7 bpm occurring at the early morning hour of 05:29 was identified as Anchor 1. The actual value is of course subject-specific and relatively low compared to other studies, however the time of occurrence is within the sleeping probability of greater than 90% and very much in line with other studies showing it to occur mid sleep [21][32]. Anchor 2, interval of maximum increase in HR, was defined as the hour before and hour after time of maximum increase in HR. From the circadian HR profile derivative signal, this ended up being defines as an increase of 7.0 bpm over the morning hours 07:42 and 09:42 when the subject is commonly waking up. From the sleep portability analysis, it is inferred that over these 2 hours the probability of the subject sleeping was becoming less and less, starting from 90% or less and ending at 25% or more. These results can be explained to directly coincide with researchers reporting HR endogenously begins to elevate as an individual approaches the end of nocturnal sleep and then surges, along with blood pressure, at the onset of awakening [16][21]. For the subject of this study, sleep-wake schedule was not consistent in timing or duration, as the weekly plots of **Figure 9** display. Nonetheless, the sleep probability gradient defends that this interval is precisely when the subject is most likely to transition from sleep to wake. Two local maximums were identified as Anchors 3 and 4, respectively; one of 64.1 bpm just after noon detected at 12:51 and the second of 67.0 bpm approximately six hours later at 18:58. Again, this confirms earlier studies concluding HR endogenously elevates in the morning and stays elevated until late afternoon [21][32]. As for which anchor is the absolute maximum, most studies identified it to be the earlier afternoon maximum, however these studies were either under constant routine protocols or USW procedure, and there was still a secondary maxima towards the later afternoon. The difference between the two local maximums detected in this study is 2.9 bpm, and between the absolute maximum and minimum it is 18.3 bpm. Therefore, in consideration of the maximum-minimum range, the two local maximums can arguably be considered as one prolonged period of elevated HR. Notably, the SE around the minimum HR anchor is detected to be but 0.2 bpm whereas at the local maximums it is 2.3 and 2.5 bpm, respectively. This can mainly be explained due to the greater variation in

activities that an individual goes through during wakeful hours of the day that elevate or depress HR from underlying endogenous values. However, HR nadir almost always occurs when the individual is sleeping and not performing any activity and therefore more precisely records HR values without other exogenous contributions besides for sleep.

A model characterizing the subject-specific circadian fluctuations of HR in free-living conditions was finally summarized and presented in **Equation 7** after evaluating different models reported in **Table 5**. The conclusion, based on the best combination of lowest number of coefficients needed to be calculated empirically, highest adjusted R^2, and lowest RMSE, a 3rd order Fourier Series model is proposed. Essentially a combination of 3 cosine and sine function pairs, this model is proposed as a mathematical equations that can characterize any individual's smoothed mean 24-hour HR profile. Interestingly, the period of the first, second and third harmonics are almost exactly 24.0, 12.0 and 8.0 hours, respectively. The absolute amplitudes of the first harmonic cosine and sine pair is largest, followed by the second, and lastly the third. Together, this model seems to perfectly capture the circadian, half circadian and quarter circadian components of a subject's 24-hour HR profile.

One of the applications interested by this research project's industry partner, Thinkinetics Inc., is being able to map and predict the circadian rhythm of an individual's HR a few days into the future. As discussed in the Introduction and Background sections of this chapter, on one hand many diseases and disorders have been linked to the desynchronization of individual's activities with respect to their circadian clock, and on the other hand there are well established links with optimal timing for best sleep, cognitive awareness, physical performance and other benefits with the individual's personal circadian rhythm [21]. For example, optimal sleeping hours are based on hormonal secretion levels which vary amongst individuals based on their specific circadian rhythm [21]. In short, there are several reported benefits to knowing and being in-tuned with one's circadian rhythm, and linking known activities with consequences – good or bad – is attainable given the advancements in wearable and physiological sensing technology, which could prove to be appreciated by clinicians and general consumers alike. To illustrate this application, **Figure 17** presents the circadian HR profile computed over all available data *except* Week 1. This profile was used to interpolate and map the profile over and along Week 1 when HR data was *not* available,

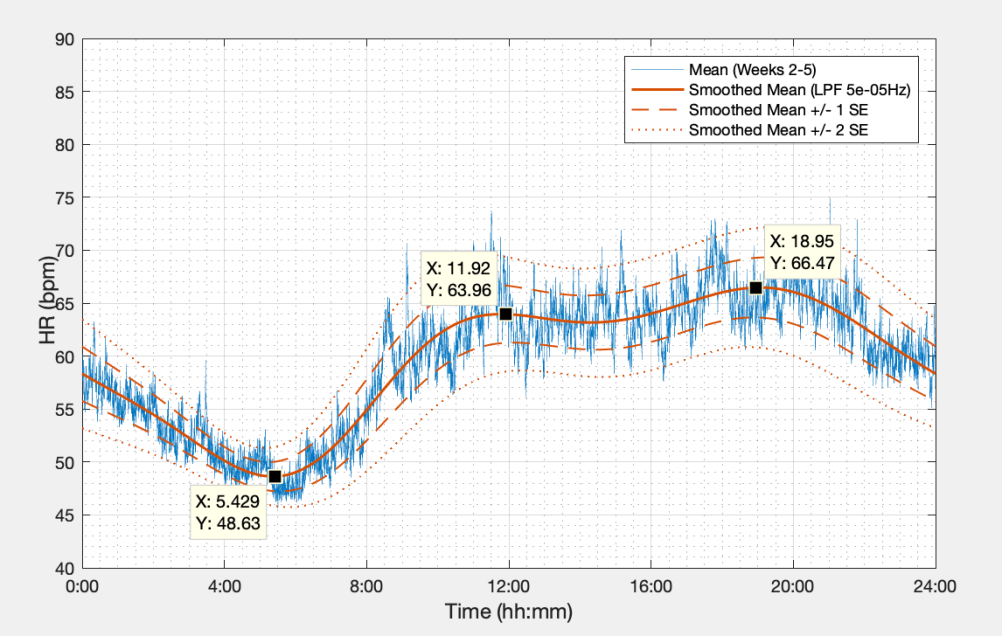

Figure 17: 24-hour mean, smoothed mean and SE boundaries on Polar HR data collected over Weeks 2 to 5. Datatips at anchors 1, 3 and 4.

as shown in **Figure 18**. The profile in **Figure 17** is highly comparable with that of **Figure 15.C** and suggests week-to-week consistency of HR circadian fluctuations. To begin with, Anchors 1, 3 and 4 are highlighted with their time and values; Anchor 1 and 4 occur at almost the exact same clock time and vary from the full dataset's' anchor values by less than 1 bpm. Anchor 3 however is almost one hour advanced and is more than 1 bpm lower in value. Analyzing **Figure 18**, it is evident that smoothed HR data majority of the time fits well within the ± 2 SE boundaries except for during the interval of exercise, upon waking up in the morning on the fourth day starting at 96 h, and to a lesser degree the short nap on that same day. Along with sleep tracking and activity tracking capabilities most smartwatches, have this type of information identifying the time of significant overshooting or undershooting of the individual's HR from a prior established circadian HR profile can lend to user decisions towards improved lifestyle behaviours and scheduling optimal with their personal circadian rhythm. Simple examples can include when to start getting ready for sleep and how late in the afternoon will exercise not delay sleep, but can even reach advising at risk individuals to take medication at times aligned with their chronobiology or immediately notify a doctor if the rate of change in morning HR is outside regular ranges.

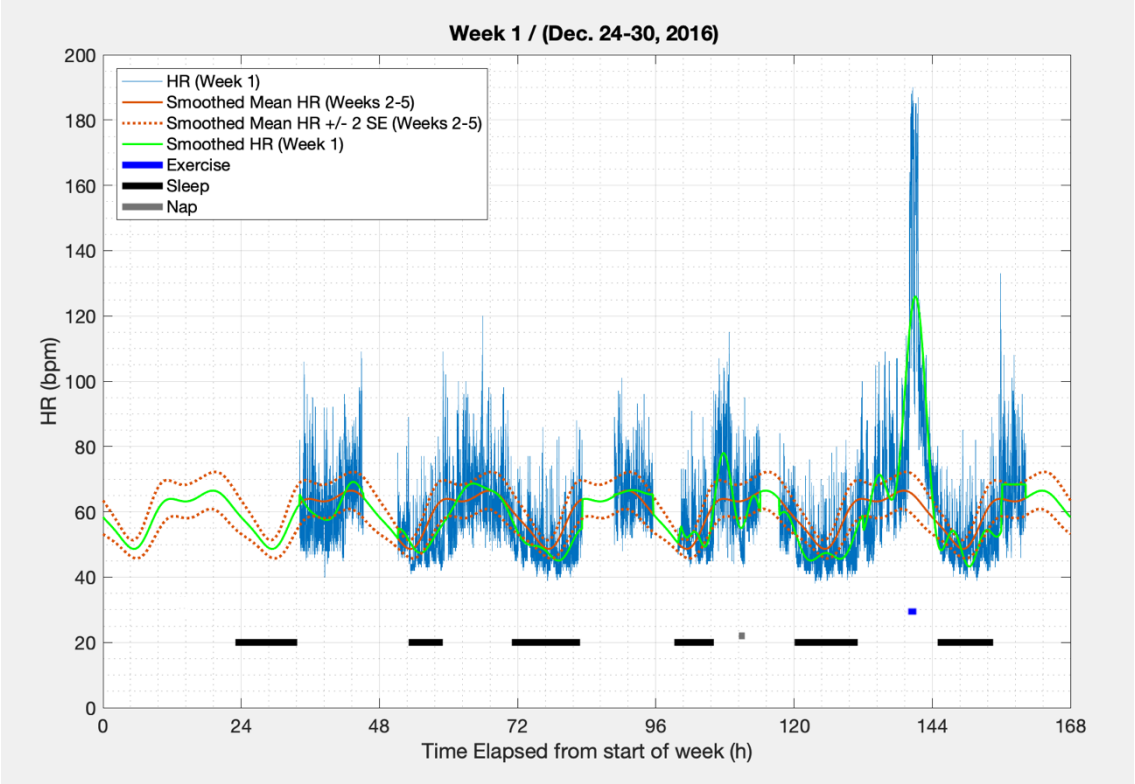

Figure 18: Week 1 Polar HR data and smoothed HR data interpolated with the smoothed 24-hour mean HR profile (computed from Weeks 2 to 5) where data is not available. External factors marked by bar lines. Smoothed mean ± 2 x SE marked as boundaries.

Chapter 3

Research Part II

3 Real-time ambulatory ECG processing and novel SQI for R-peak-derived cardiac vital signs

3.1 Introduction

Electrocardiography, practically invented by Willem Einthoven in 1895, is the process of recording the electrical activity of the heart by placing electrodes on the skin producing an electrocardiogram signal (process and signal abbreviated as either ECG or EKG). The micro voltage changes from each cardiac depolarization-repolarization cycle is detected by the electrodes and produces characteristic waveforms in the ECG signal. ECG today is part of the gold standard for evaluating general health and has become a very powerful tool for detecting, diagnosing and monitoring an array of cardiovascular diseases (CVD), which accounted for 31% of all deaths worldwide in the year 2016 [35]. Historically, ECG has been monitored under stationary conditions, usually in a supine position, and in hospitals or clinics. However, in the 1950s, Norman Holter invented a portable ECG recorder, what is called today a Holter monitor. The Holter monitor lasts 24 to 48 hours and meant to be sent home with patients known or being tested for CVD in order to collect continuous, uninterrupted monitoring later to be analyzed by clinicians and cardiologists [36]. Since then, plenty of remote-monitoring ECG devices have been released in the market for research and as certified medical devices after passing rigorous standards set out by governing bodies such as the Food and Drug Administration (FDA) in the United States and Health Canada in Canada.

The challenge though with monitoring ECG outside the hospital in ambulatory environments is that the ECG inherently becomes very susceptible to many noise sources. One in particular – electrode motion artifact – is known to be the toughest to handle since it presents itself as ectopic beats or as high power white noise; in both cases making it very difficult to analyze the underlying physiological signal [37]. Therefore, mitigating against electrode motion artifact is highly sought out by many researchers. From a hardware perspective, electrode material science (e.g. flexible capacitive electrodes [38]) and electrode-skin contact (e.g. dry versus gel [39]) are

but two means investigated that can limit the thresholds and recurrences of motion artifact corrupting the signal. From a signal processing perspective, creative ways passed simple filters are continuously being proposed as a means to extract underlying cardiac vital signs despite these type of artifacts. Adaptive filtering, machine learning models and sensor fusion when available are some of the options proposed in the literature to accurately identify heart beats and other ECG information in the presence of noise. Nonetheless, at one point or another, motion can be so great that noise will inevitably be introduced and render entire segments unreadable. At this point, such segments needs to be identified as unreliable for ECG analysis so that inaccurate vital signs are not being outputted and false events that would otherwise caution a potential health crisis are not being triggered.

3.1.1 Research Objectives and Scope

The main objective of this research is to develop a real-time algorithm which processes ambulatory ECG data and outputs reliable cardiac vital signs HR and HRV. In order to achieve this, first an R-peak detection algorithm needed to be developed with attention towards mitigating noise factors common in ambulatory ECG data. Second, an efficient and accurate HR and HRV calculation method was employed on the R-peak detection results. And lastly, a SQI on the underlying ECG segment used to calculate HR and HRV was developed and proposed to accompany each cardiac vital sign as a quantifiable measure to how reliably these cardiac vital signs should be considered.

The scope of this project involved developing a real-time ECG processing algorithm, as outlined above, in MATLAB language such that it can be easily translated and employed in the software of a wearable ECG-monitoring medical device. Specifically, the developed algorithm was to be used by this project's industry partner, Elastic Care Inc., who were developing a medical-grade remote patient monitoring device with 12-lead ECG monitoring along with respiration and movement. Ambulatory ECG data was not collected as a part of this study, therefore the algorithm was developed and evaluated on publicly available ambulatory ECG databases: the Massachusetts Institute of Technology - Beth Israel Hospital (MIT-BIH) Arrhythmia database and the MIT-BIH Noise Stress Test (NST) database [37]. Performance metrics were calculated and assessed based on what has been brought down in literature and based on the applicable American National Standard ANSI/AAMI/IEC 60601-2-47:2012/(R)2016, which is required to be followed by

ambulatory ECG systems submitting for FDA or Health Canada approval as an ECG monitoring medical device [40].

3.2 Background

The ECG signal is a valuable diagnostic tool since with each beat of the heart a characteristic waveform is produced, which specialists are able to interpret in regards to cardiovascular health and diseases. **Figure 19** illustrates the characteristic P-QRS-T waveform for normal heartbeats, termed sinus rhythm. Minor differences in the characteristic sinus rhythm to the trained eye is a specific medical indication. Different ECG leads, achieved by placing unipolar or bipolar electrodes at different locations along a subject's chest, allows for an even deeper layer of analysis as to which atrium or which ventricle is the source of irregular cardiac electrical activity; twelve leads is standard for a complete ECG analysis. A non-extensive list of main applications for reasons to monitor an individual's ECG include: detection of irregular heart rhythms (also referred to as arrhythmias), coronary artery blockage, areas of damaged heart muscle, enlargement of the heart, inflammation of pericarditis, electrolyte imbalance, lung diseases, monitoring heart medication and pacemakers, monitoring ongoing heart attacks, and ruling out hidden heart diseases in pre-surgery assessment [29].

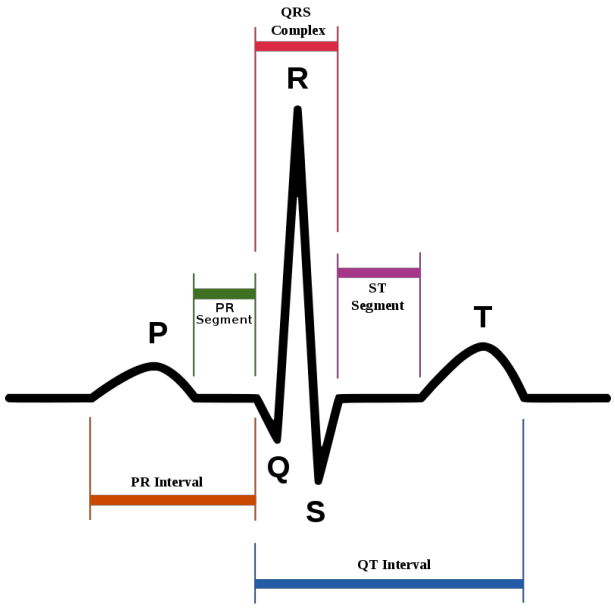

Figure 19: Schematic diagram of the characteristic P-QRS-T waveform of sinus rhythm in humans as seen in the ECG signal. [34]

The QRS complex, namely the R-peak, is generally accepted as the reference time of each heartbeat. One of the most well-known real-time QRS detection algorithms is the one developed by Jiapu Pan and Willis J. Tompkins in the year 1985 [41]. The algorithm, recognized as the Pan-Tompkins Algorithm, is divided into three process: learning phase 1, learning phase 2 and detection. Learning phase 1 requires about 2 seconds so that at least one R-peak presents itself and initializes R-peak and noise amplitude thresholds based on the assumption that the largest peak is R-peak and other peaks at least half in ECG amplitude are noise peaks (or T-waves). Learning phase 2 requires two heartbeats and is in order to initialize RR interval and RR rate thresholds. The detection process consists of sequential processing steps including three linear digital filters and adaptive thresholding. The first steps involve a low pass filter (LPF) followed by a high pass filter (HPF), essentially achieving a bandpass filter, in order to reject noise components outside the fundamental frequencies of the QRS complex. The design goal of the bandpass filter was to only keep frequency components in the range of 5-15 Hz, however they achieved 3 dB passband approximately between 5-12 Hz due to processing limitations. Following the bandpass filter, the derivative of the resulting signal was approximated by a filter and then a simple point-by-point

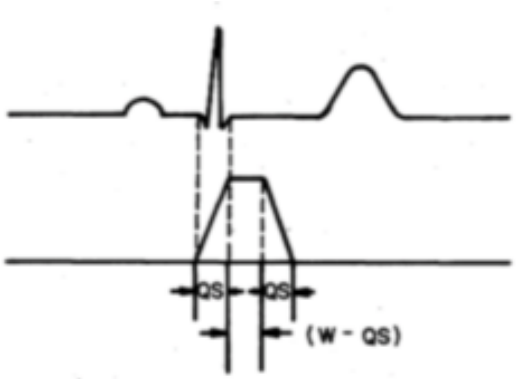

Figure 20: Idealized illustrative comparison of input ECG signal (top) and output after the Moving-Window Integration step (bottom), highlighting to relationship of the QRS complex width, *QS*, and the integrator window width, *W*. [41]

squaring operation. Next, a 150 ms moving-window integration was performed (the authors explain that the 150 ms width of the window was determined empirically). **Figure 20** illustrates the results on the steps thus far explained starting from input ECG signal to after the moving-window integration.

The step that came next was beat-by-beat detection of QRS complexes, termed fiducial mark by the authors. As can be deduced from **Figure 20**, this was achieved by detecting the point of maximal slop on the first rising edge of the integration output, which corresponds to the R-peak. The last two steps after this involved threshold testing and updating; one based on amplitude of peak detected and the other based on the resulting interval between the peak detected and the previous peak. The amplitude threshold essentially classifies each detected peak as either an R-peak or a noise peak by first testing whether the amplitude at the point of detection on the integration signal is closer to the previously detected R-peak or previously detected noise peak, and second if the amplitude of the point of detection on the bandpass filtered signal is also closer to the previously detected R-peak or previously detected noise peak. Integration signal and bandpass filtered R-peak amplitudes, respectively, along with those of each tested and failed peak (i.e. noise peak) are updated with each detection. After an acceptable R-peak is determined from the first thresholding step, an average of the eight most recent RR intervals are computed and compared against a Low, High and Missed threshold limit based on the average of the eight most recent RR intervals that fell between the Low and High limit thresholds. The Low threshold is essentially 92% of the average RR interval, the High limit 116% and the Missed limit 166%. If the new detected R-peak renders the average to be above the Missed threshold, the maximal peak reserved that will establish the average RR interval to be between the Low and High thresholds is reconsidered to be a true R-peak. This concludes the steps and process of the Pan-Tompkins Algorithm and it should be noted that all threshold percentages were empirically determined.

The Pan-Tompkins Algorithm was evaluated on all 48 records of the MIT-BIH Arrhythmia database with performance results presented in **Table 6.** Total number of beats, FP beats, FN beats, and failed detection beats and rates are the record-by-record performance metrics of choice and in summary their algorithm correctly detects 99.3% of QRS complexes on this standard dataset of ambulatory, pathological ECGs.

Since the Pan-Tompkins Algorithm has been published, several modifications have been proposed, including the one by Patrick S. Hamilton and W. J. Tompkins himself where they improved the peak detection step, optimized a set of QRS peak and noise thresholds, and removed the average RR interval threshold (which limited the Pan-Tompkins Algorithm to ECG records of normal heart rates, performing poorly on bigeminy and trigeminy arrhythmias for example) [42]. The Hamilton-Tompkins Algorithm reported 99.69% and 99.77% gross sensitivity (Se) and gross

positive predictivity (Pp) on the MIT-BIH Arrhythmia database. Another famous and competing class of R-peak detection algorithms was first developed by Zong *et al.* in 2003 and involves a length transform operation instead of just digital filters [43]. Zong *et al.* reported gross Se and Pp of 99.65% and 99.77%, respectively, on the standard MIT-BIH Arrhythmia database.

Table 6: Results of evaluating J. Pan and W. J. Tompkin's Real-Time QRS Detection Algorithm on the MIT-BIH Arrhythmia database [41].

Tape (No.)	Total (No. Beats)	FP (Beats)	FN (Beats)	Failed Detection (Beats)	Failed Detection (%)
100	2273	0	0	0	0
101	1865	5	3	8	0.43
102	2187	0	0	0	0
103	2084	0	0	0	0
104	2230	1	0	1	0.04
105	2572	67	22	89	3.46
106	2027	5	2	7	0.05
107	2137	0	2	2	0.09
108	1763	199	22	221	12.54
109	2532	0	1	1	0.04
111	2124	1	0	1	0.05
112	2539	0	1	1	0.04
113	1795	0	0	0	0
114	1879	3	17	20	1.06
115	1953	0	0	0	0
116	2412	3	22	25	1.04
117	1535	1	1	2	0.13
118	2275	1	0	1	0.04
119	1987	1	0	1	0.05
121	1863	4	7	11	0.59
122	2476	1	1	2	0.08
123	1518	0	0	0	0
124	1619	0	0	0	0
200	2601	6	3	9	0.35
201	1963	0	10	10	0.51
202	2136	0	4	4	0.19
203	2982	53	30	83	2.78
205	2656	0	2	2	0.08
207	1862	4	4	8	0.43
208	2956	4	14	18	0.60
209	3004	3	0	3	0.10
210	2647	2	8	10	0.38
212	2748	0	0	0	0
213	3251	1	2	3	0.09
214	2262	2	4	6	0.26
215	3363	0	1	1	0.03
217	2208	4	6	10	0.45
219	2154	0	0	0	0
220	2048	0	0	0	0
221	2427	2	0	0	0.08
222	2484	101	81	182	7.33
223	2605	1	0	1	0.04
228	2053	25	5	30	1.46
230	2256	1	0	1	0.04
231	1886	0	0	0	0
232	1780	6	1	7	0.39
233	3079	0	1	1	0.03
234	2753	0	0	0	0
48 patients	116 137	507	277	784	0.675

No matter the R-peak detection algorithm, all attempt to eliminate FPs and FNs caused by ECG artifacts, however all fail at certain SNRs where the underlying signal is no longer retrievable. ECG artifacts can be characterized into three categorized: physiological, experimental and environmental [29]. The first arises due to other physiological processes in the body such as muscle artifacts from the electrical activity during muscle contractions or baseline wander from respiration. The second includes all the artifacts more prevalent or more of a concern during ambulatory ECG monitoring and include electrode-motion artifact and electrode-contact artifact. These artifacts can present themselves at all frequencies and at powers significantly greater than the electrical activity of the heart, thus mimicking QRS complexes. The last category of ECG artifacts are those such as powerline interference, electromagnetic interference and interference from circuit components [29]. Because ECG artifacts can easily corrupt the signal from interpretability, SQIs become a valuable tool to relay if the ECG signal is clean enough for diagnostic inspection and/or cardiac vital sign extraction.

The goal of ECG SQIs are to extract signal and beat properties and output binary classification on individual segments of data as either being acceptable (1) or unacceptable (0). Literature discusses physiological, temporal and frequency domain SQIs. Physiological ECG SQIs are usually heuristic and evaluate RR interval and rate with respect to known physiological limits. Orphanidou *et al.* proposes HR in the range of 40 and 180 bpm at rest and 40 and 300 bpm during exercise, maximum RR interval of 3 seconds (which allows for one missed beat if 40 bpm is physiological minimum) and a maximum RR interval divided by minimum RR interval limit in a 10 second segment of 2.2 [44],[45]. Temporal domain SQIs are both heuristic and empirical and include flat line detection, minimum and maximum amplitude limits such as 0.2 and 15 mV, respectively, amplitude range limits such as ± 4 mV, excessive amplitudes for a prolonged duration such as 2 mV for 0.2 seconds or longer, and large amplitude rate of changes such as 0.005 mV/sec or greater [46]-[49]. The third and fourth moment of signal segments (i.e. skewness and kurtosis, respectively) are other temporal domain SQIs with acceptable values determined empirically from labelled datasets [50]. Frequency domain SQIs involve segmenting the ECG into different bandwidths and computing in-bound to out-of-band or in-band to full-band power ratios, respectively, and then set empirically determined limits for acceptable ratio values [48],[51].

A study by Clifford *et al.* extracted several of these SQIs and used them as features to train different machine learning models to classify whether ECG segments are clean or noisy [52]. On the PhysioNet Challenge 2011 dataset, consisting of 12-lead ECG records with binary signal quality labels for each lead, they segmented the data into 5 second segments and achieved classification testing accuracies of 98% and 97%, respectively, from multi-layer perceptron and support vector machine (SVM) models.

3.3 Methods

3.3.1 ECG Database

Two publicly available ECG databases from PhysioNet were used in this study: (1) the MIT-BIH Arrhythmia database, and (2) the MIT-BIH NST database [37]. The MIT-BIH Arrhythmia database consists of 48 half-hour excerpts of two-channel ambulatory ECG recordings digitized at 360 Hz. Only modified limb lead II (MLII) were used of the two channels and records with paced beats were excluded, resulting in 44 records used from this database. The labels of MLII records used are the following: 100, 101, 103, 105, 106, 108, 109, 111, 112, 113, 114, 116, 117, 118, 119, 121, 122, 123, 124, 200, 201, 202, 203, 205, 207, 208, 209, 210, 213, 214, 215, 220, 221, 223, 231, 233, 234.

The MIT-BIH NST database, consisting of 12 half-hour records, uses two patient records from the Arrhythmia database (records 118 and 119) and adds a combination of three types of artificially generated noise most common to ambulatory ECG at six particular SNRs of -6, 0, 6, 12, 18 and 24 dB, respectively. The three artificial noise signals generated were designed to mimic baseline wander, muscle artifact and electrode motion artifact. The combined noise signal was introduced to the records after the first 5 minutes of recording and thereafter for 2 minute segments alternating on and off every 2 minutes. MLII for all 12 NST records were used: 118e_6, 119e_6, 118e00, 119e00, 118e06, 119e06, 118e12, 119e12, 118e18, 119e18, 118e24, and 119e24.

The NST database reflects different extremes of noise able to be introduced from being recorded in an ambulatory environments, and the Arrhythmia database, recorded in actual ambulatory environments, reflects the different types of pathological patients that can distort otherwise sinus ECG signals and still must be processed as reliable when noise is or is not present. Together, majority of the signal properties an ECG monitoring device would encounter in the

ambulatory environment are considered. The Arrhythmia database was used to evaluate performance of the overall algorithm developed and the NST database was used as a tool to extract signal properties at different noise levels contributing to unreliable cardiac vital signs.

3.3.2 Algorithm Development

3.3.2.1 Real-time processing

The real-time ECG processing algorithm was designed to have an initial baseline data collection period followed by 1-second sliding windows where signal processing operations (1) R-peak detection, (2) HR and HRV calculation, and (3) SQI calculation are performed on each window until recording halts.

The baseline data collection period is intended to extract baseline signal properties and initial cardiac vital signs under the assumption no motion artifact was present. The user pairing with the ambulatory ECG monitor would be instructed to remain still for this baseline duration – essentially an initial calibration which is not uncommon in remote monitoring devices – and then baseline parameters such as amplitude thresholds, significant frequency components and initial cardiac vital signs can be extracted. The specific baseline data collection outputs are listed below:

- Amplitude range (maximum minus minimum)
- Most recent 12 RR intervals
- Minimum RR interval
- Most recent HR estimate
- Minimum HR estimate
- Most recent HRV estimate

Extracting this type of baseline information is proposed to improve performance on the subsequent signal processing steps in the event environmental noise does get introduced downstream.

The length of each sliding window is intended to keep enough *past* ECG data in order to most accurately detect most recent R-peaks and signal quality features with the trade-off of not storing too much memory to be employed by a monitoring device. Then, keeping the same window length, the window slides by 1-second and re-runs the signal processing steps in order to output results on the next 1-second of data, and so on and so forth. The 1-second sliding window

parameter was selected in order to output new cardiac vital signs at a rate of 1 Hz in real-time if implemented on a remote monitoring device. This would allow for fast response to sudden changes in cardiac activity, a feature assumed desirable to most wearable and medical cardiac monitoring devices.

3.3.2.2 R-peak detection algorithm

A modified Pan-Tompkins Algorithm was implemented for this project's R-peak detection algorithm. Namely, the LPF, HPF, derivative filter and point-by-point squaring operations were implemented as Pan and Tompkins reported in their paper. The moving-window integration was implemented as a moving averaging filter and the fiducial mark maximum slope detection followed by adaptive amplitude and RR interval thresholding was switched to a peak detection function with minimum peak prominence threshold, adaptive minimum peak-to-peak distance threshold and an optimization search. The steps of the algorithm developed are listed below with descriptions of each operation implemented sequentially, starting from input ECG:

Step 1: LPF

Convolution with transfer function (MATLAB's *'conv'* function) presented in **Equation 8**:

$$H(z) = \frac{(1 - z^{-6})^2}{(1 - z^{-1})^2} \tag{8}$$

Step 2: HPF

Convolution with transfer function presented in **Equation 9**:

$$H(z) = \frac{(-1 + 32z^{-16} + z^{-32})}{(1 + z^{-1})} \tag{9}$$

Result of the LPF and HPF steps is a 3 dB passband approximately between 5-12 Hz [41].

Step 3: Derivative

Convolution with transfer function presented in **Equation 10**

$$H(z) = (f_s/8)(-z^{-2} - 2z^{-1} + 2z^{1} + z^{2}) \tag{10}$$

where f_s is the sampling rate, equal to 360 Hz for the MIT-BIH Arrhythmia and NST databases.

Step 4: Squaring

Simple point-by-point squaring operation.

Step 5: Averaging

Convolution by the transfer function presented in **Equation 10**:

$$H(z) = \frac{1}{L}\sum_{k=0}^{L-1} z^{-k} \tag{10}$$

where L is length in sample number of the moving average window size, which is set to 150 ms, therefore $L = 0.15f_s = 0.15(360) = 54$ for the databases used.

Step 6: Peak detection with Adaptive Thresholds

MATLAB's *'findpeaks'* function with 'MinPeakProminence' and 'MinPeakDistance' specified. 'MinPeakProminence' set to 5% of maximum peak in segment. 'MinPeakDistance' initialized to 200 ms (corresponding to an instantaneous HR of 300 bpm) for the baseline data collection period and then updated with each sliding window as 55% of the shortest RR interval from the 12 most recent RR intervals. Thresholds were determined empirically on the dataset.

Step 7: Peak location Optimization

Location of the peaks detected on the moving average signal were optimized to the location on the input ECG signal within ± 20 ms that resulted in the maximum squared value. This optimization was identified visually and then empirically after testing searching window length from ± 10 to 60 ms. Searching for maximum value or maximum squared values was also compared and maximum squared value (essentially maximum absolute value) yielded better performance, which can be explained due to the arrhythmias which some produce negative voltage R-peaks.

3.3.2.3 HR and HRV calculations

HR is calculated as the equal-weighted average of the N most previous RR intervals, converts it into units of beats per minute (bpm), and outputs second-by-second updated values with each signal processing window. The HRV calculation method was designed to simply output the

most recent RR interval value in units of milliseconds (ms) after each signal processing window. **Equations 11 and 12** presents the HR and HRV calculation equations, respectively:

$$HR_i = \frac{60}{\frac{1}{N}\left(RR_i + RR_{i-1} + RR_{i-2} + \cdots + RR_{i-(N-1)}\right)} = [\text{bpm}] \tag{11}$$

$$HRV_i = 1000 \times RR_i = [\text{ms}] \tag{12}$$

where RR_i is the most recent RR interval, in units of seconds, of the i^{th} signal processing window, and HR_i and HRV_i are similarly the HR and HRV outputs, respectively, from the i^{th} signal processing window.

As default, N was set to equal 12 for the HR calculation equation, however in cases where each of the most recent three RR intervals are greater than 1200 ms (representing instantaneous HRs less than 50 bpm) the value of N switched to equal 4. It is common to adjust the number of RR intervals averaged based on their lengths when approximating real-time HR for medical purposes. A similar decision rule is reported in GE Healthcare's CARESCAPE Patient Data Module [53]. In practice, N is commonly set to either 8 or 12. The greater the value of N, the slower it responds to sudden increases or decreases in HR however also outputs less variation between successive HR outputs. Since HRV output represents individual RR-interval variation, HR calculation was decided to average over 12 instead of 8 RR-intervals. Together, the HR and HRV outputs captures the gradual changes and immediate variation of heart rate, respectively.

It should be noted that, as **Equations 11 and 12** show, accuracy of HR and HRV are 100% dependent on the accuracy of the R-peak detection algorithm. As explained, the number of RR intervals used to approximate the cardiac vital sign only affects over how many heartbeats is the vital sign value captured over; however, the R-peak detection performance determines if the vital sign is reliable or not regardless of the approximation chosen.

3.3.2.4 Novel SQI for R-peak derived cardiac vital signs

Since noise can inevitably be assumed to corrupt ambulatory ECG recordings, this research set out to develop a novel SQI which quantifies the reliability of R-peak derived cardiac vital signs HR and HRV. As such, each second-by-second output is accompanied by a real number SQI

ranging from 0 to 1, where at the lower limit 0 indicates a completely unreliable underlying ECG segment and at the upper limit 1 indicates a completely reliable segment. This SQI aims to suggest with what degree of certainty the cardiac vital sign should be considered reliable and derived from noise-free ECG.

The SQI proposed computes 14 well-established signal features, derives 9 decision rules based on previously reported thresholds and from training a support vector machine (SVM) on the NST database, and then weighs the respective contribution of each decision rule from a Random Forest (RF) classification analysis to output the final SQI. **Table 7** presents the 14 features (F1-14) and 9 decision rules (DR1-9) used in developing the proposed SQI along with their source.

Table 7: Features and Decision Rules used to develop the novel ambulatory ECG-derived cardiac vital sign SQI.

Feature	Feature Value	Decision Rule	Decision Rule for "Passed" (i.e. DR = 0)	Source
F1	Heart rate	DR1	$35 \leq F1 < 180$ bpm (rest) $40 < F1 < 300$ bpm (active)	[45]
F2	Max. RR interval	DR2	$F2 \leq 3$ sec	[44]
F3	Max/min. RR interval	DR3	$F3 \leq 2.2$	[44]
F4	Duration of longest Flat line	DR4	$F4 \leq 0.2$ sec	[46]
F5	Max.−min. Amplitude	DR5	$F5 \leq 4$ mV or 120% max. baseline	[48]
F6	Duration of amplitude above 2 mV	DR6	$F6 < 0.2$ sec	[49]
F7	No. of consecutive *steep* slopes	DR7	$F7 \geq 1$	[49]
F8	Power in 0.05-0.25 Hz band	DR8	SVM Classifier on F8-13 (trained on the MIT-BIH NST database)	[48]
F9	Power in 0.25-10 Hz band			
F10	Power in 10-20 Hz band			
F11	Power in 20-58 Hz band			
F12	Power in 58-62 Hz band			
F13	Power in 62-100 Hz band			
F14	Spectral density ratio 5-14/5-50 Hz	DR9	$0.40 < F14 < 0.95$	[51]

All features are taken from or extracted over the 1-second segment of signal processing windows. For F1, HR is calculated as defined in **Equation 11**. If there are no R-peaks detected in the 1-second output segment, F1 still assumes the most recent segment's HR value, however F2 will equal 0 and F3 is set to one. If there is only one R-peak detected, F2 can be calculated as the

interval between the detected R-peak and the previous one but F3 equals 1, and if there are two or more R-peaks detected, F2 is the most recent RR interval and F3 considers all RR intervals involved in the 1-second output segment. For F4, flat line segments are defined as segments where the signal slope equals zero. For F7, *steep* slope is defined as per [49] as a rate of change greater than +0.0005 mV/sec. Power in specifically defined bands F8-13 are common ECG bands separating low, ECG1, ECG2, medium, powerline and high frequency bands, respectively. The spectral density ratio (SDR) also considers well-defined ECG bands where the power in 5-14 Hz isolates QRS complex power and is then considered with respect to the power of the overall predominant band of ECG signals.

Decision rules were taken from the literature and five were slightly modified as results were being investigated on specific records of this study's dataset:

Firstly, since the dataset used in this study is that of pathological patients, some noise-free record segments displayed the arrhythmia bradycardia resulting in HRs below 40 bpm at rest when considering DR1. The lower limit resting HR set in [45] was changed to 35 bpm as default, however if within the baseline data collection period a lower HR was calculated, that value less 5 bpm was chosen as the lower limit. Ideally, HR limits should be an operator-selectable parameter set by a patient's physician or the user themselves, which the developed algorithm can easily accommodate for.

Secondly, the amplitude range can somewhat vary based on ECG recorder, ECG lead and/or electrode. The dataset used in this study appeared to produce different amplitude ranges than the ones reported in [48] with several clean records surpassing the 4 mV threshold. Therefore, in cases when 4 mV amplitude range was surpassed, DR5 was empirically modified to allow up to 120% of the initial baseline period's peak amplitude range before considering the segment to have failed this decision rule.

Thirdly, *steep* slopes did arise sometimes in clean segments of data, assumingly due to specific arrhythmia conditions, therefore DR7 was modified to allow up to one *steep* slope.

Fourthly, SVM classification on solely power in different frequency bands presented in DR8 was not found in literature; for example, [48] also included F5 as a feature input. The SVM classifier developed in this study used only F8, F9, F10, F11, F12 and F13 as input features. Furthermore, the outputs of training and testing datasets commonly used are labeled as either clean or unclean with respect to signal quality annotated by a cardiologist or ECG expert. The dataset

used to develop this study's SVM classifier was the full MIT-BIH NST database and the novel binary classification output selected was passed (0) or failed (1) R-peak detection results on the 1-second output segment the features were extracted over. Essentially, each record went through the full real-time ECG processing algorithm with second-by-second R-peaks detected and features extracted. R-peak detection performance was then evaluated by comparing its detection accuracy with the annotated R-peaks and each 1-second output with one or more false detections were assigned a label of "failed" and each with no false detections were assigned a label of "passed". As such, the features and labels were defined, each feature-label pair randomized and then first split into an 80:20 train:test dataset to test different SVM parameters on the test dataset. Once the highest performing parameters were identified, the full dataset was used to train the final SVM classifier as DR8.

And lastly, the SDR reported by [51] on the MIMIC II database assigned the acceptable range to be between 0.50 and 0.80, which this study did not confirm. Empirically iterated to achieve highest SQI performance on the NST database, the acceptable range for DR9 was set to be between 0.40 and 0.95.

Once all features were extracted and decision rules defined and empirically optimised, this study set out to combine the results of the decision rules (i.e. count of which passed and which failed) not equally, but rather by assigning greater weight to the more important features/rules. By presenting **Equations 13 and 14**, the difference between an equally-weighted and a weighted combination of decision rules to arrive at a single, real number SQI value in the range of [0,1], respectively, is demonstrated:

$$SQI = \frac{DR_1 + DR_2 + \cdots + DR_N}{N} \tag{13}$$

$$SQI = \frac{a_1 DR_1 + a_2 DR_2 + \cdots + a_N DR_N}{\sum_{n=1}^{N} a_n} \tag{14}$$

N is the number of decision rules considered in the calculation, DR_n is the n^{th} decision rule which either equals 1 (pass) or 0 (fail), and a_n, in **Equation 14** only, represents the weight assigned to the n^{th} decision rule.

To assign appropriate weights, a RF classifier developed by Kyriaki Kostoglou *et al.* in [55] was utilized. In the training phase, the RF model is designed to consider features randomly selected and then iteratively replaced on different subsets of the dataset. The left out data from the subset, termed the out-of-bag (OOB) set, is used for validation as the decision tree grows and generates the OOB error. The average OOB error across all observations of different feature combinations is then used to identify most important features as those that produce the least OOB error [55]. Similarly as the SVM classifier, the binary classification used to train the RF classifier was R-peak detection "pass" or "fail" on 1-second output segments of the MIT-BIH NST database. Once the RF classifier was defined on this dataset, best performing feature set and their relative performance was used to approximate an integer multiple to their corresponding decision rules as outlined in **Table 7**.

3.3.3 Analysis

3.3.3.1 R-peak detection, HR and HRV analysis

R-peak detection performance is evaluated for each record of the databases individually and then overall for each of the two databases used, respectively. The metrics reported for each record are TP, FP and FN number of beats, and statistics Se and Pp which are calculated as shown in **Equations 15 and 16**, respectively.

$$Se = \frac{TP}{TP + FN} \tag{15}$$

$$Pp = \frac{TP}{TP + FP} \tag{16}$$

Importantly to note, in order to consider a detection a TP the detection was allowed to be within ± 150ms of the annotated peak label, as per standard [40] and commonly reported in literature for R-peak detection performance.

For the respective databases, two aggregate statistics are commonly reported: (1) gross statistics, which assigns equal weight to each detection, and (2) average statistics, which assign each record equal weight. These record-by-record and aggregate database statistics are what is requested to be reported for ECG monitors filling for medical device status. [40]

Record-by-record vital signs HR and HRV performance are evaluated by each of their RMSE with reference vital sign computed in the same manner as presented in **Equations 11 and 12** just with RR intervals calculated from the labeled beats, as instructed in standard [40]. The RMSE calculation formula is presented in **Equation 17**.

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(E_i[n] - R_i[n])^2} \tag{17}$$

where E_i is the estimated cardiac vital sign over the i^{th} segment based on the R-peak detection algorithm results, R_i is the reference cardiac vital sign over the i^{th} segment based on the annotated beat labels, and M is the number of segments (i.e. the duration) of the record. $RMSE$ is computed on HR and HRV, respectively, for each record individually and only the average statistic is computed for each respective database's performance.

3.3.3.2 SQI analysis

SQIs have been commonly studied as binary classifications to differentiate between poor/unacceptable and good/acceptable signal quality [45][48][56]. However, when SQIs are combined, some studies do represent them in the range of 0 to 1 or 0% to 100%. In these studies, the objective is to decide which ECG lead or which physiological signal (e.g. arterial blood pressure) should be used to estimate HR based on the signal that generates the highest SQI [51][56]. However in this study, the scope is individual ECG lead analysis and the motivation is to output a real-number SQI indicating *how* reliable the underlying signal is in estimating the cardiac vital sign. Therefore, it is clear that evaluating the performance of this study's proposed SQI is how well it differentiates between ECG segments yielding wrong cardiac vital signs, and hence failed R-peak detections, and to what extent are these wrong or failed, respectively. A high performing SQI model would intuitively assign lower SQI values to ECG segments yielding large HR and HRV error and low R-peak detection accuracies, and successively increasing SQI values as the HR and HRV errors decrease and R-peak detection accuracies increase. Conversely, a poor SQI model would yield no linear relationship between SQI value and cardiac vital sign error and R-peak detection performance.

As such, each group of 1-second output ECG segments were grouped according to same SQI value outputs and analyzed with respect to the same R-peak detection, HR and HRV performance metrics presented earlier: R-peak Se and Pp, HR RMSE, and HRV RMSE.

3.4 Results

The design of the real-time algorithm developed is presented in **Figure 21**. As can be seen, after an initial baseline data collection period of 20.5 seconds the first HR and HRV vital signs are outputted with an automatic initial SQI of 1.0 (since assumed noise-free). Thereafter, these outputs – HR, HRV and SQI – are updated with each second of recording that continues. The signal processing sliding windows were set to 10.0 seconds in order to balance between keeping enough

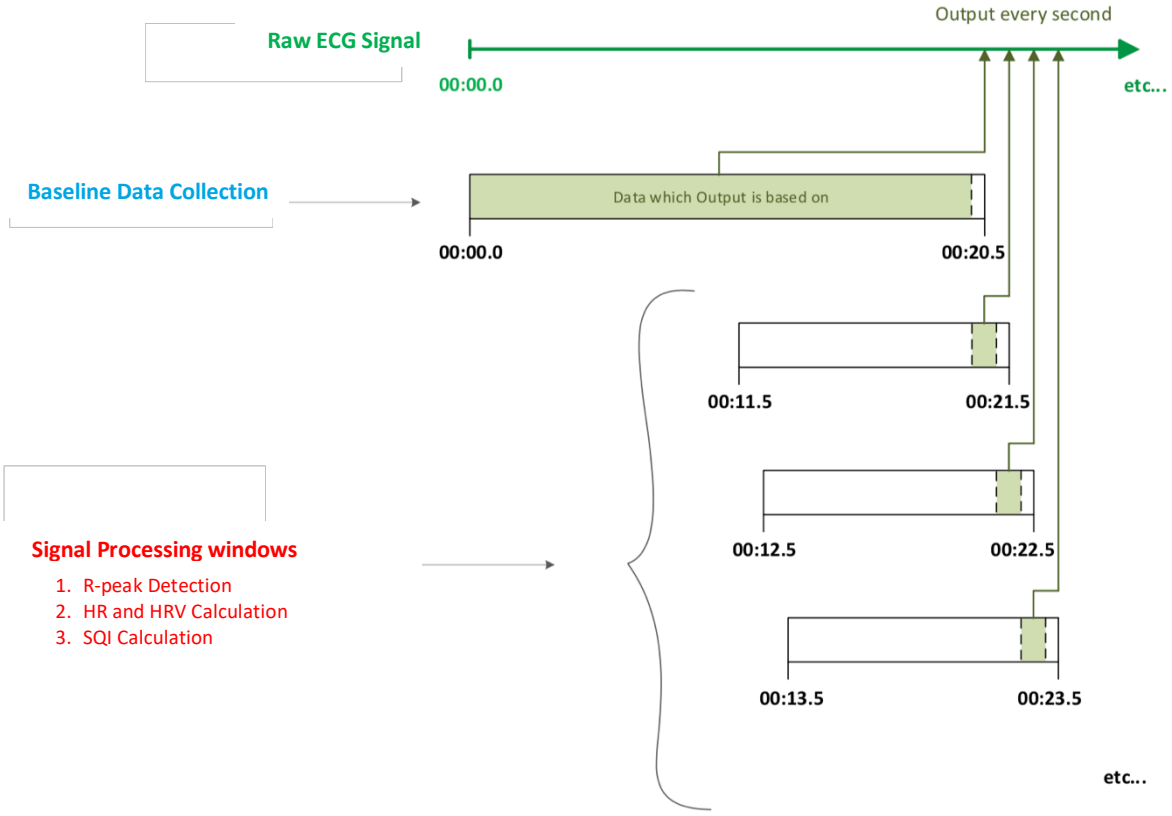

Figure 21: Design of the Real-time ECG Processing Algorithm developed. The segments of the raw ECG signal used in baseline data collection and then in sliding-window signal processing is represented by the signal timeline (green arrow; units in mm:ss of elapsed recording time): baseline data collection spans the first 20.5 seconds of recording, which then outputs baseline signal properties and initial outputs on the first 20.0 seconds of data. Thereafter, signal processing windows 10 seconds in length slide by 1.0 second and performs (1) R-peak detection, (2) HR and HRV calculation, and (3) SQI calculation on the full window and outputs the new results calculated over the 8.5th to 9.5th second of the respective 10 second window, thus outputting new second-by-second updated cardiac vital signs and reliability measure.

data for high R-peak detection performance and acceptable memory if employed by a monitoring device. The last 0.5 seconds of each window, baseline data collection period included, are discarded in the event the window ends in the middle of a QRS complex which has been identified to result in false detections. Therefore, there is a delay of 0.5 seconds (not including processing time delays) between the time of output and the physiological time of occurrence.

For illustration purposes, **Figure 22** presents the processed ECG output with detected R-peaks and second-by-second SQI values overlaid in the background (using the same y-axis) on a selected record – record 118 of the MIT-BIH Arrhythmia database.

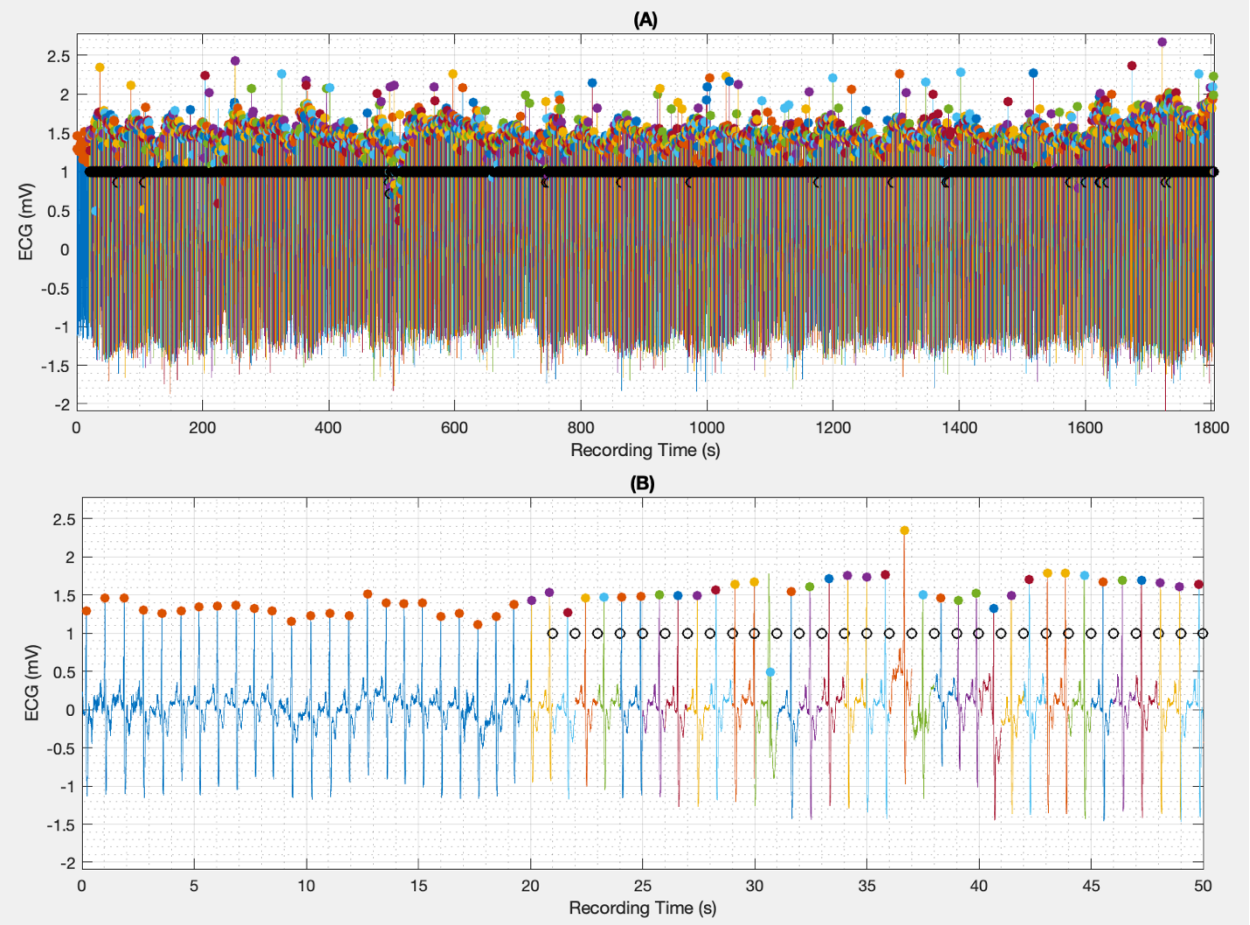

Figure 22: R-peak detection and SQI output results on MIT-BIH record 118; **(A)** Complete record 1804 seconds long of processed ECG data; **(B)** Zoomed-in section of the first 50 seconds of recording. The initial blue ECG trace and red filled circles R-peaks over the 0^{th} to 20^{th} second are the outputs from the baseline data collection period. Thereafter, alternating colours of trace and filled circles are the outputs from each 1-second segment output of the sliding window signal processing steps. At the end of each sliding window signal processing step, a black circle (ranging from 0.0 to 1.0) represents the SQI output of that segment.

On a segment in the same record selected above, step-by-step outputs from the modified Pan-Tompkins algorithm developed is presented in **Figure 23**.

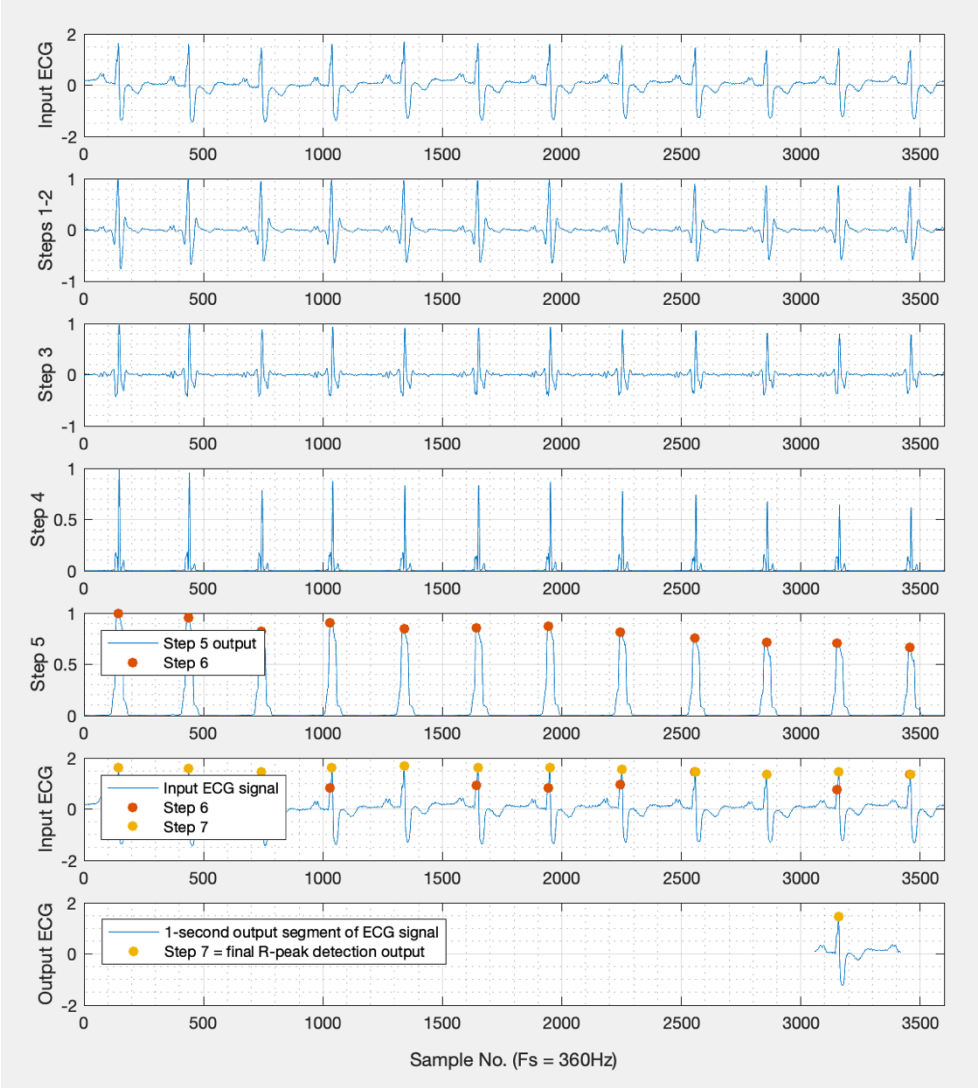

Figure 23: Step-by-step results of the modified Pan-Tompkins algorithm developed and employed on the 63rd to 64th 1-second output segment of MIT-BIH record 118. Starting with 10 second input ECG (in units of mV and sampled at a rate of 360Hz), the steps go through the bandpass filter (Steps 1-2), the derivative filter (Step 3), squaring (Step 4), moving-window average (Step 5), peak detection the averaged signal (Step 6), optimizing the peak detection on the ECG signal (Step 7) and the only outputting the R-peaks detected on the 1-second output segment which is between the 8th and 9th second of the 10 second window.

Table 8 and Table 9 on the next pages present final R-peak detection, HR and HRV performance statistics on the MIT-BIH Arrhythmia and NST databases, respectively.

Table 8: R-peak detection, HR and HRV performance statistics on the MIT-BIH Arrhythmia database.

Record	R-peak Detection Statistics					HR RMSE (bpm)	HRV RMSE (ms)
	TP	FP	FN	Se	Pp		
100	2270	0	0	1	1	0.13	14.29
101	1863	2	1	0.9995	0.9989	0.39	19.19
103	2082	0	0	1	1	0.09	1.77
105	2568	33	2	0.9992	0.9873	6.09	51.80
106	2011	1	15	0.9926	0.9995	2.00	76.03
108	1742	191	19	0.9892	0.9012	20.90	194.96
109	2529	0	1	0.9996	1	0.49	16.21
111	2120	0	2	0.9991	1	0.64	29.76
112	2537	0	0	1	1	0.13	2.87
113	1791	0	2	0.9989	1	0.56	23.87
114	1864	1	13	0.9931	0.9995	1.44	98.43
115	1951	0	0	1	1	0.08	2.59
116	2387	3	22	0.9909	0.9987	3.57	217.51
117	1533	0	0	1	1	0.26	11.96
118	2276	1	1	0.9996	0.9996	0.72	22.08
119	1985	0	0	1	1	0.16	15.04
121	1861	0	0	1	1	0.12	4.23
122	2474	0	0	1	1	0.11	2.44
123	1514	0	3	0.9980	1	0.73	33.82
124	1617	0	0	1	1	0.11	15.75
200	2595	23	4	0.9985	0.9912	3.30	52.93
201	1951	0	11	0.9944	1	1.88	74.44
202	2122	0	12	0.9944	1	2.17	71.71
203	2958	53	20	0.9933	0.9824	6.58	70.05
205	2639	0	15	0.9943	1	4.72	50.18
207	1849	418	9	0.9952	0.8156	36.65	705.14
208	2928	8	25	0.9915	0.9973	4.84	160.48
209	3003	6	0	1	0.9980	3.27	15.50
210	2637	13	10	0.9962	0.9951	4.14	50.03
212	2746	0	0	1	1	0.12	2.27
213	3248	0	0	1	1	0.24	9.39
214	2246	3	14	0.9938	0.9987	2.48	80.93
215	3355	0	5	0.9985	1	1.20	26.30
219	2152	0	0	1	1	0.12	7.90
220	2041	0	5	0.9976	1	1.01	29.43
221	2422	0	3	0.9988	1	0.83	27.07
222	2477	8	4	0.9984	0.9968	2.80	49.99
223	2602	0	1	0.9996	1	0.49	25.05
228	2048	67	4	0.9981	0.9683	8.94	110.70
230	2254	0	0	1	1	0.21	9.45
231	1563	0	6	0.9962	1	1.34	81.25
232	1780	8	0	1	0.9955	1.19	72.82
233	3073	0	3	0.9990	1	0.95	30.50
234	2750	0	1	0.9996	1	0.47	12.11
Total	**100414**	**839**	**233**	**-**	**-**	**-**	**-**
Average	**-**	**-**	**-**	**0.9977**	**0.9914**	**2.92**	**60.91**
Gross	**-**	**-**	**-**	**0.9977**	**0.9917**	**-**	**-**

Table 9: R-peak detection, HR and HRV performance statistics on the MIT-BIH NST database.

Record	R-peak Detection Statistics					HR RMSE (bpm)	HRV RMSE (ms)
	TP	FP	FN	Se	Pp		
118e_6	1973	848	304	0.8665	0.6994	31.42	327.21
119e_6	1826	915	159	0.9199	0.6662	41.55	353.07
118e00	2174	780	103	0.9548	0.7360	37.32	266.52
119e00	1943	805	42	0.9788	0.7071	41.86	336.92
118e06	2254	587	23	0.9899	0.7934	32.65	251.84
119e06	1977	615	8	0.9960	0.7627	35.54	305.45
118e12	2274	234	3	0.9987	0.9067	19.22	159.21
119e12	1985	252	0	1	0.8873	18.66	208.29
118e18	2276	38	1	0.9996	0.9836	4.96	66.26
119e18	1985	12	0	1	0.9940	2.06	44.45
118e24	2276	1	1	0.9996	0.9996	0.72	22.13
119e24	1985	0	0	1	1	0.18	16.38
Total	**24928**	**5087**	**644**	**-**	**-**	**-**	**-**
Average	**-**	**-**	**-**	**0.9753**	**0.8447**	**22.18**	**196.48**
Gross	**-**	**-**	**-**	**0.9748**	**0.8305**	**-**	**-**

Results of HR and HRV outputs with respect to their accompanying SQI are presented in **Figure 24** on record 118 of the MIT-BIH Arrhythmia database, the same record selected thus far to use as example. The bottom panel of **Figure 24** represent the residual between subtracting the outputted vital sign from the reference vital sign as well as identifies the points of failed R-peak detection. In contrast, the same results for record 207 of the MIT-BIH Arrhythmia database, which is one of the records yielding worst performance as can be seen in **Table 8**, are presented in **Figure 25**. Upon inspection of **Figure 25**, at the times of several failed R-peak detections, especially FPs, the SQI begins to decrease well below unity. **Figure 26** zooms in on one of the failed intervals and also presents the underlying ECG signal, which clearly shows poor signal quality of such that the annotators could not retrieve the beat rhythm or location (hence no annotated beats).

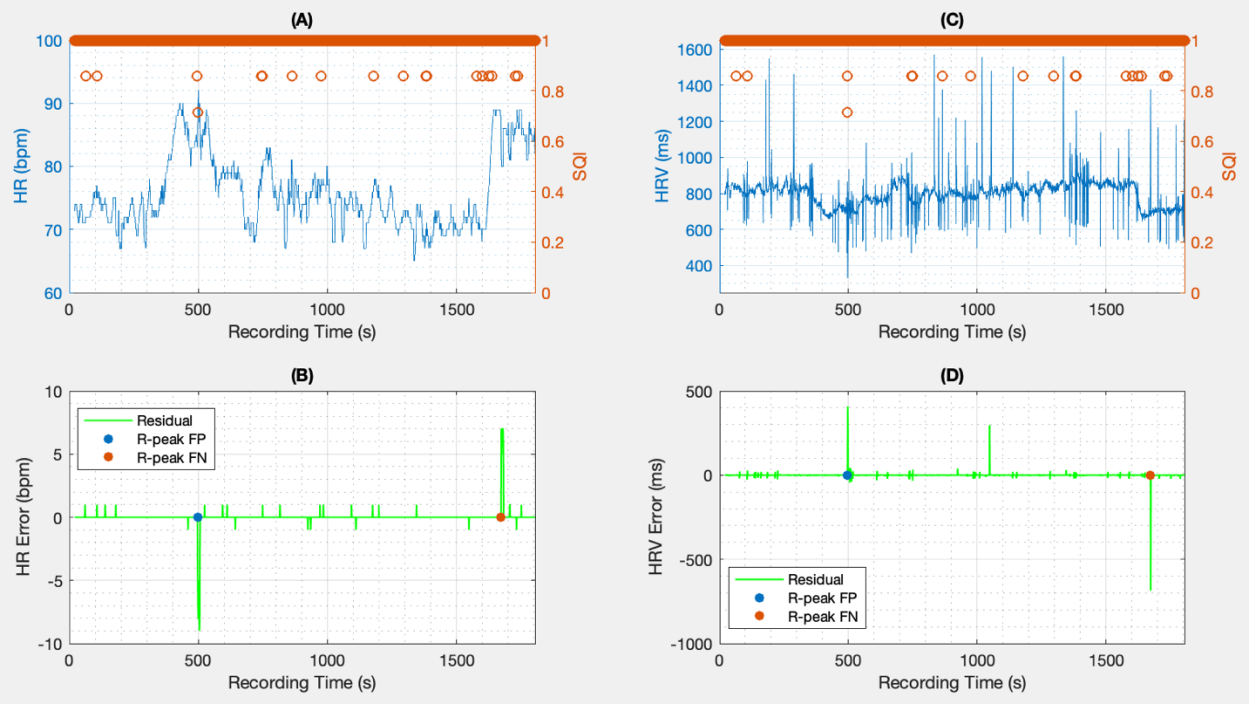

Figure 24: HR **(A)** and HRV **(C)** second-by-second output results on MIT-BIH record 118, along with each output's accompanying SQI. HR Error **(B)** and HRV Error **(D)** with respect to estimates using the labelled R-peaks, highlighting the time at which the R-peak detection algorithm failed.

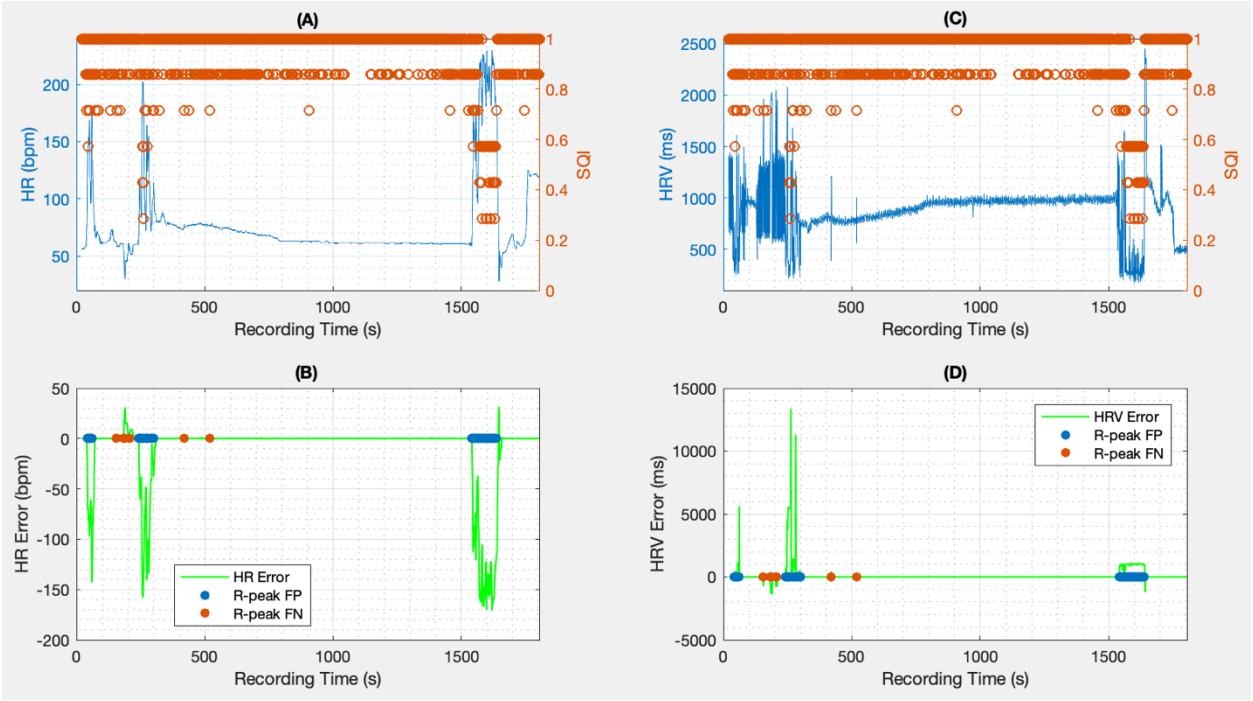

Figure 25: HR **(A)** and HRV **(C)** second-by-second output results on MIT-BIH record 207, along with each output's accompanying SQI. HR Error **(B)** and HRV Error **(D)** with respect to estimates using the labelled R-peaks, highlighting the time at which the R-peak detection algorithm failed.

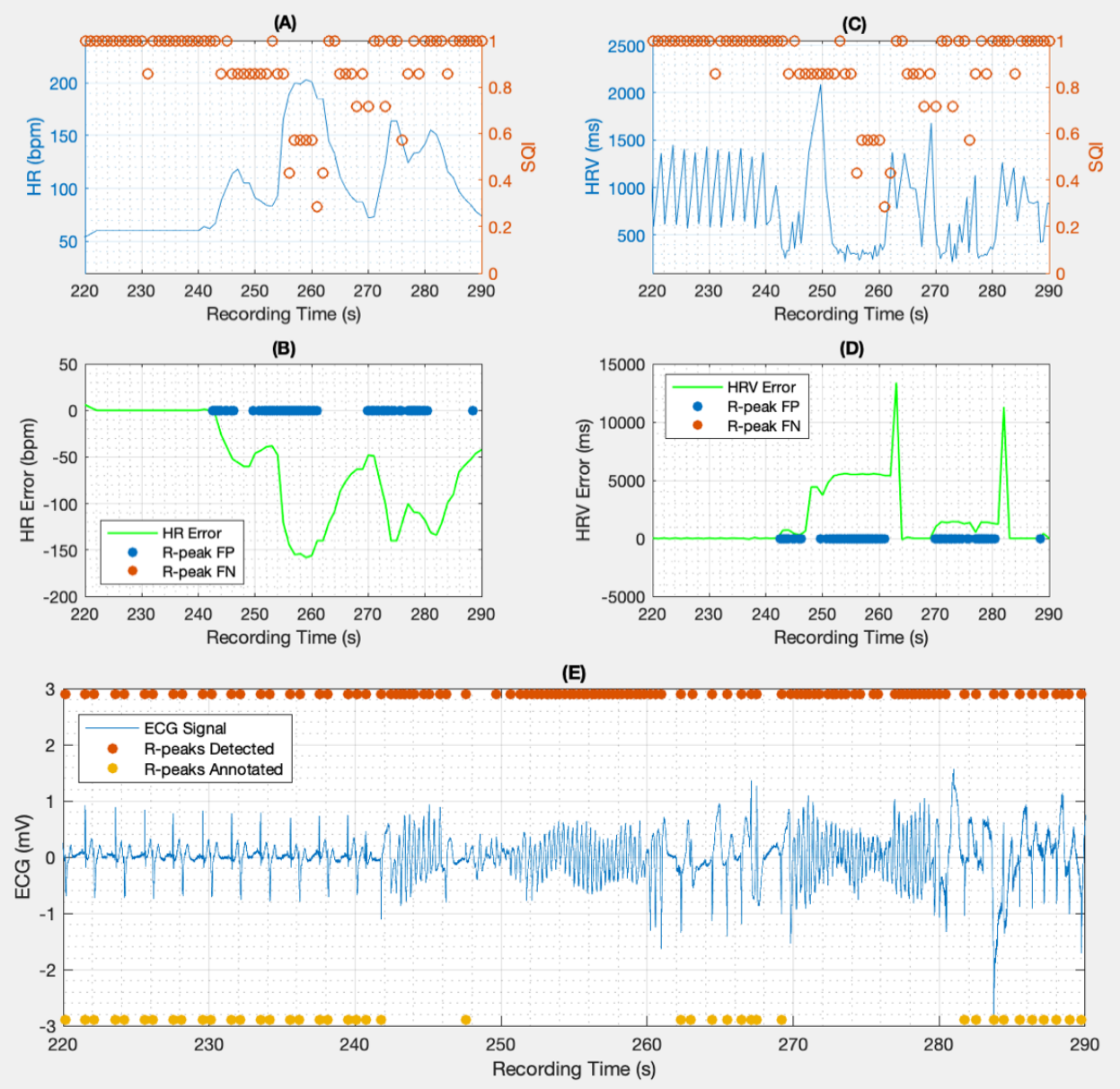

Figure 26: MIT-BIH Arrhythmia record 207; **(A-D)** Zoomed-in section of **Figure 25**; **(E)** Output's corresponding ECG signal with the algorithms detected beats (red, top) and the annotation labeled beats (yellow, bottom).

The final SQI model proposed was presented in **Equation 14**. The weights (a_n) of each decision rule (DR_n) concluded from this study are presented below:
(Recall $DR_n = 1$ if "passed", and 0 if "failed")

Equation 14 weight assignments for proposed SQI:

- **i.** Default: all $a_n = 1$
- **ii.** If $DR_2 = 1$: $a_1 = 3$
- **iii.** If DR_2 passed and DR_3 failed: $a_2 = 2$ and $a_3 = 1$;
  If DR_2 failed and DR_3 passed: $a_2 = 1$ and $a_3 = 2$;
- **iv.** If $DR_4 = 0$: $a_4 = 0$
- **v.** If $DR_6 = 0$: $a_6 = 0$

These assignments were made from the RF classifier analysis after investigating most important features and their relative differences with adjustments made empirically from the SQI performance analysis on the MIT-BIH Arrhythmia database. The feature importance results from the RF classifier trained on R-peak detection performance on the MIT-BIH NST database are presented in **Figure 27**.

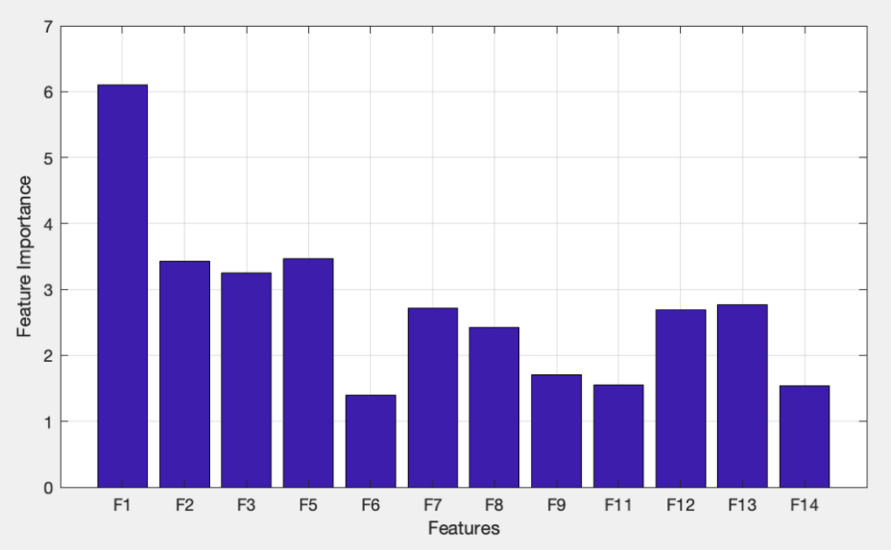

Figure 27: Important features of highest performing RF Classifier trained on the MIT-BIH NST database using R-peak detection performance ("pass" or "fail") as the binary classification labels.

In order to interpret SQI performance on the Arrhythmia database, 1-second segments were grouped based on SQI output values and then R-peak detection statistics, HR RMSE and HRV RMSE computed and analyzed. **Table 10** and **Figure 28** presents these results for the case when SQI was derived from equally weighting all decision rules (i.e. **Equation 13**) and from weighting the decision rules as defined above for the final proposed model.

Table 10: Comparison of SQI results computed from the final proposed model of weighted decision rules (DRs) and equally-weighted DRs, respectively, on all 1-second segments of ECG data from the MIT-BIH Arrhythmia database processed through the real-time algorithm developed.

(Equation 14) SQI = Weighted DRs based on RF Classifier				**(Equation 13)** SQI = Equal Weighted DRs			
SQI Groups	Total no. segments	No. Passed segments	% Passed segments	SQI Groups	Total no. segments	No. Passed segments	% Passed segments
0.2857	12	3	0.2500	0.5556	2	2	1.0000
0.4286	37	13	0.3514	0.6667	66	52	0.7879
0.5714	170	94	0.5529	0.7778	1196	1132	0.9465
0.7143	1696	1523	0.8980	0.8889	8982	8867	0.9872
0.8571	8420	8331	0.9893	1	68250	67963	0.9958
1	68250	67963	0.9958				
Total	**78496**	**78016**	**680**	**Total**	**78496**	**78016**	**680**

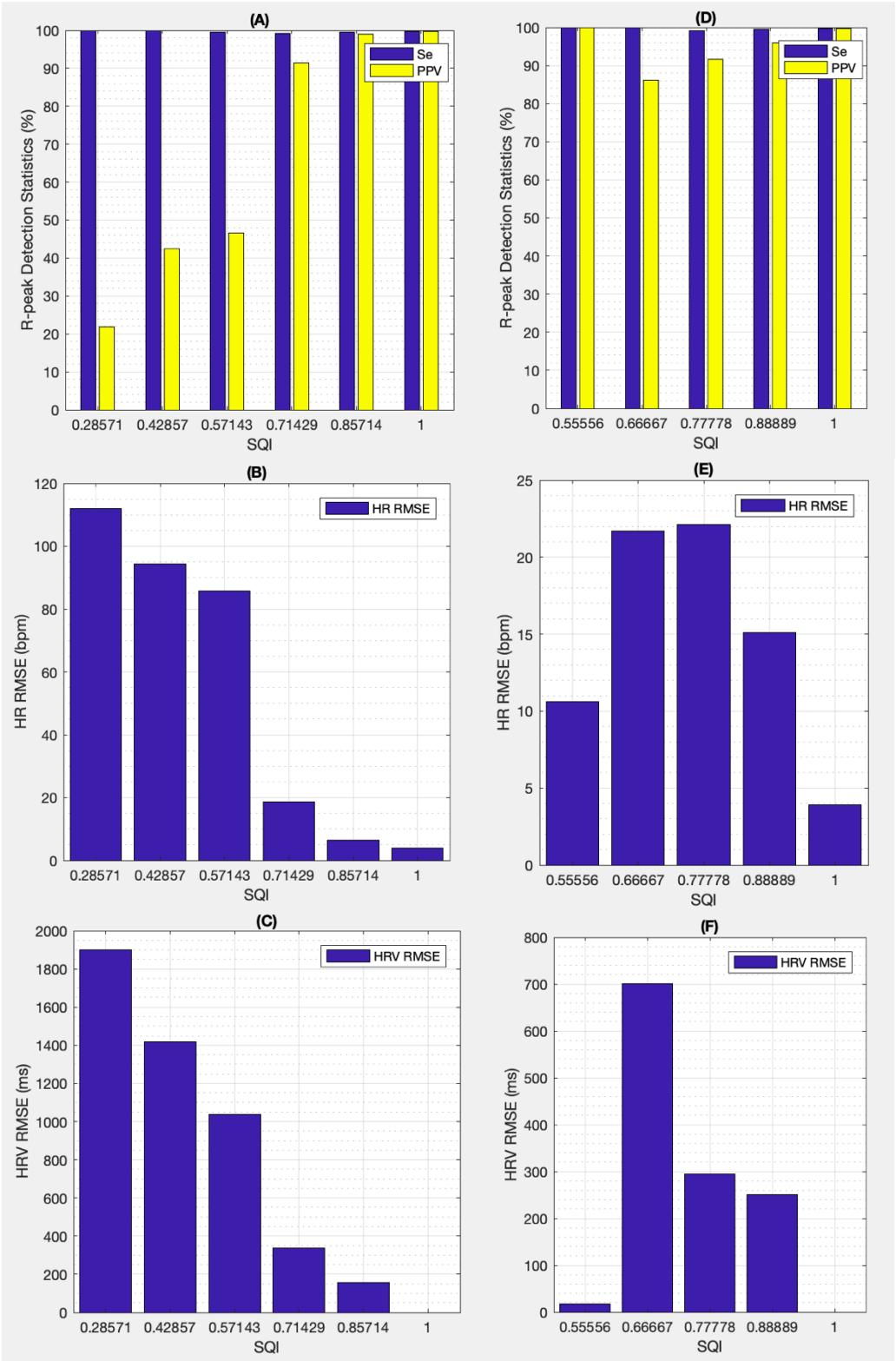

Figure 28: Results of SQI performance analysis; **(A-C)** SQI calculated from the final proposed weighted average of DRs; **(D-F)** SQI calculated from equally-weighted DRs.

3.5 Discussion

The overall ECG algorithm developed was successfully designed to accommodate real-time processing through an initial baseline data collection period followed by sliding windows. It is not uncommon for wearable or medical devices to instruct an initial calibration period before commencing a new recording session and can therefore be used to set initial user-specific thresholds and extract noise-free signal properties. Second-by-second updates for HR and HRV along with a novel SQI quantifying how reliable these R-peak derived cardiac vital signs are were successfully outputted.

The R-peak detection algorithm is at the core of how accurate cardiac vital signs – HR and HRV – are when outputted. Since HR can be estimated over different lengths of time or different number of RR intervals, it is the detection accuracy of QRS complexes which determines reliability and the estimation method but needs to be reported. Same applies for HRV, which usually outputs RR interval values on a beat-by-beat basis as oppose to second-by-second most recent RR intervals as done in this project.

A modified Pan-Tompkins R-peak detection algorithm was implemented and yielded high performance comparable to prior studies reporting on the classic MIT-BIH Arrhythmia database. As presented in **Table 8**, 99.77% gross Se and 99.17% gross Pp were achieved, which is slightly more sensitive than the results reported for the Hamilton-Tompkins algorithm at the expense of being a little less precise [42]. There is a natural balance between how sensitive an R-peak detection algorithm can be to not miss any detections and how precise it can be to get everyone. This was seen when testing different parameters and thresholds and what rules were considered for them to update with each window of processing. This led to empirically iterating through different values, visually inspecting results and looking for highest performance on the overall MIT-BIH NST database with discrete segments of different SNRs and on select records with annotated segments indicating poor signal quality from the MIT-BIH Arrhythmia database. Outputs after each step of the R-peak detection algorithm and what segment of the window was used for most recent cardiac vital estimation were presented in **Figure 23**. This figure clearly shows how the last step of a search window for picking the time of the R-peak precisely and not anywhere else in the QRS complex was successfully achieved. This final optimization step minimizes minor variations and maintains consistency across R-peak derived cardiac vital sign

estimates. Also reported in **Table 8**, is the performance of HR and HRV, which yielded 2.92 bpm RMSE and 60.91 ms RMSE, respectively, as an average across all records on the MIT-BIH Arrhythmia database. In context of the normal range for HR and HRV, respectively, these errors are minor and appear acceptable.

Arriving at the final SQI for quantifying reliability of HR and HRV outputs involved a few steps combining methods brought down in literature in a novel way based on this project's specific goal. Well-established features and decision rules for their acceptable limits or range were extracted on 1-second ECG segments. Features were physiological-, temporal- and frequency-based and for the most part the decision rules were empirically determined by several researchers. For this reason, some of the decision rules were changed to meet the properties of this study's dataset and goals. A SVM classifier, trained on the MIT-BIH NST database used power in six ECG bands as features and "passed" or "failed" R-peak detection results as classification labels. This was used to output an independent decision of whether or not the ECG segment is reliable or not to output cardiac vital signs based on frequency content distribution. Once all individual decision rules were defined, a RF classifier was similarly trained on the NST database, this time with the purpose of identifying important features contributing to each decision rule. **Figure 27** illustrates the RF feature importance results of the highest performing model, which as can be seen, only consists of 12 of the 14 records: features 4 (F4) and 10 (F10) are excluded. Referring back to **Table 7,** F4 extracts flat line information from the ECG segment, however it was identified that not one record in both databases used in this study had a flat line duration greater than the limit set in the corresponding decision rule. Therefore, it makes sense that this feature was dropped from the best performing model and essentially removed from one of the decision rules considered in the final SQI model – i.e. a_4 assigned to equal zero in the final model for **Equation 4**. Similarly, F6, which is the feature that extracts excessive amplitudes above a threshold, was also rarely encountered in the dataset. So although F6 is not dropped from the final feature list, it holds the lowest feature importance and therefore its decision rule was also assigned a weight equal to 0 if not failed. The other feature not present in the final feature list, F10, is one of the ECG bands and interestingly enough is the ECG2 band of 10 to 20 Hz where some of the QRS complex frequency content exists. The two highest bands were most significant, which makes sense since true noisy segments resulting in failed R-peak detection results usually consist of high frequency noise.

Together, the six power band features (F8-13) were considered to be moderately important and therefore their unique decision rule (the SVM classifier) maintained a weight of 1. Most evident in **Figure 27** is that F1, the physiological limit of HR, is the most important feature in distinguishing passed or failed R-peak detections. This does not come as a surprise since HR estimate accuracy is completely dependent on R-peak detection accuracy. For equal reasons, next most important features are F2 and F3 which extract RR interval thresholds. At first glance, it appears that F1 is about four times more important than the least important features, F1 and F2 are about double, F7 and F8 are somewhere around one and a half, and F12 and F13 are part of the SVM classifier decision rule. Therefore, this information helped guide the initial weight assignment which was then iteratively corrected to the proposed one which yielded results highly suggesting a well characterized SQI for R-peak detection performance and cardiac vital sign reliability.

In order to identify if the RF classifier feature importance did add value to the SQI model proposed, **Table 10** presented count results for how many ECG segments were assigned to unique SQI values computed without and with the input from the RF classifier analysis; namely, equally weighting each decision rule as if uninformed which ones contain more information about R-peak detection performance, and the proposed SQI model, respectively. As can be seen from these results, the proposed model not only distinguishes SQI values more uniformly between the full range of 0 and 1, but also yields much better passed versus failed or % passed segments than the equal weighted SQI. the equal-weighted SQI model only outputs 4 unique SQI values which do not even range the full 0 to 1 scale the reliability measure is intended to cover. Even more reassuring is that the % passed segments for the proposed SQI model are very comparable to the actual corresponding SQI values, whereas for the equally-weighted SQI model the % passed segments do not match the SQI values for the full range at all. **Figure 28** demonstrates this moreover by presenting the performance of each group of SQIs with respect to R-peak Se and Pp, HR RMSE and HRV RMSE. Most strikingly from these results is how the proposed SQI differentiates increasingly improved R-peak detection precision (i.e. Pp statistic) with an approximate linear trend, and similarly HR and HRV error linearly decreases. Conversely, the equal weighted SQI results show no correlation with performance and SQI value over the full range of unique SQI values.

Further analyzing the performance results of the proposed SQI model, very interestingly is how at SQI equal to 1 the HRV RMSE is nil but there is a relatively significant HR RMSE. This can be explained due to HRV being calculated based on the two most recent R-peak detections (most recent RR interval) and HR is essentially average based on intervals over the 13 most recent R-peak detections. Hence, despite the present ECG segment not containing any noise resulting in failed R-peak detections, the HR estimate outputted may still contain error from prior segments that were corrupted with noise. This presents itself as a delay limitation, however an inherent limitation to any average HR estimate as oppose to an instantaneous HR measurement such as the HRV calculated in this study.

Coming back to training the initial classifiers and identifying important features which differentiate between ECG segments that are more likely to yield failed R-peak detection results or not, I would like to reemphasize that this was exactly the goal of this study rather than the commonly used low/poor/unacceptable segment versus high/good/acceptable segment binary classifications. The proposed SQI is not intended to suggest whether or not the ECG signal is legible or of high enough diagnostic quality, it is intended to provide quantifiable information to how accurate the derived cardiac vital signs are on the segment of ECG data used. It is also noteworthy to mention that the method executed for training and defining the proposed SQI is agnostic to the R-peak detection algorithm upstream the overall algorithm. Meaning, if another R-peak detection algorithm is used or the current one improved, the same features and decision rule results can be extracted, fed into the RF R-peak detection pass or fail classifier to determine important features on a dataset of your choosing and then analyzed to appropriately weight each decision rule to arrive at the final combined SQI. This is an attractive feature which can allow for algorithm updates, even remotely (via remote Bluetooth firmware updates for example), to improve performance of the SQI rather than hardware and signal integrity improvements, which may not always be economically feasible or practical if the device is already in the field.

Figure 29 presents an illustration to help explain the main application of the proposed SQI: With each update in HR and HRV, an accompanying score – the proposed SQI – will update as well indicating to what degree of reliability the vital signs given to you should be considered with. As a biomedical engineer working in the field of wearable devices, this is a common concern of users: “How can I trust the outputs I am getting?”.

Figure 29: Illustration of the application for this study: with each cardiac vital sign outputted to the users an SQI quantifying how reliable these estimates are is provided. Background image taken from [57].

Chapter 4
Conclusions

4 Conclusions

4.1 Key Findings and Relevance

The studies performed in this thesis involved both HR signal and ECG signal analysis, respectively, collected from wearable devices. The former was a direct output collected from an individual subject paired with consumer wearable devices, and the latter was taken from an open source database of medical-grade ambulatory recordings. The main goal of this thesis was thus to reliably collected and interpret biological information from these datasets in spite of the ambulatory condition common to wearable device monitoring.

In the first part of this research, an individual's diurnal, endogenous fluctuations of HR was investigated in free-living conditions. Circadian rhythm studies are not commonly performed in uncontrolled environments since (a) monitoring devices outside the laboratory have many limitations including battery life and susceptibility to ambulatory artifacts, and (b) there are several external factors contributing towards masking the endogenous, diurnal fluctuations. This study thus set out to address both of these concerns. To begin with, the PPG-derived HR from the smartwatch was compared to a smartshirt with ECG-derived HR and was concluded to be as reliable and even less susceptible to high frequency changes arising from ambulatory artifacts at times. Next, from recordings collected over about five weeks total, the diurnal endogenous fluctuations was explored despite several data-collection interruptions and free-living conditions. A 24-hour profile of the endogenous HR fluctuations was successfully separated from the available data then smoothed with a LPF to emphasize times and HR values of key anchors reported in literature. From there, the smoothed circadian HR signal was parametrized with a 3rd order Fourier Series as a proposal for an individualized model characterizing endogenous fluctuations in HR. Furthermore, the sleep-wake schedule was creatively mapped over the 24 hours of a day as an individual's probability of sleeping at any given hour. As such, the circadian HR profile and key anchors were given context with respect to the sleep-wake schedule, which is one of the most significant and unavoidable external factor contributing to circadian rhythm entrainment.

In the second part of this research, a real-time ambulatory ECG processing algorithm with HR and HRV outputs was developed. The real-time design involved an initial baseline data collection period followed by sliding windows of continuous processing until recording terminates. Over each window of data, a modified Pan-Tompkins algorithm detecting R-peaks for HR and HRV estimates was implemented and shown to yield high and comparable performance to the prior art on a standard ambulatory ECG database of pathological subjects. Since ambulatory ECG is highly susceptible to motion artifacts, a SQI relaying how reliable each cardiac vital sign output is was proposed. Starting with common signal features with reported or empirically determined decision rules separating clean and noisy ECG segments, a RF classifier determined which features and corresponding decision rules were most important in distinguishing segments with failed R-peak detection results. Based on relative feature importance from the RF analysis, results from each decision rule were combined to generate a novel SQI quantifying how reliable each ECG segment is. With this type of information accompanying each HR and HRV value outputted, the user can better evaluate for themselves how accurate the data they are receiving is at any given moment.

In summary, the findings from this thesis show promise that there are signal processing techniques, creative algorithms and specific analyses that can be employed on biological signals collected from wearable devices in order to output reliable information.

4.2 Future Research

Since wearable devices are capable of monitoring multiple physiological parameters and are continuously adding new ones to their suit of features, sensor fusion and the co-dependency of biological signals and processes is naturally a future area of interest which can lend to both advancements in free-living circadian rhythm studies and improved vital sign extraction methods. After identifying the diurnal, endogenous fluctuations of HR, with each new physiological parameter added, such as CBT, respiration or sweat analytes, and which is then modeled as a subject-specific 24-hour profile, a deeper understanding of how our internal biological clock plays a role in our timing of activities and behaviours can further be studied. As for the SQI proposed, it can only relay if the individual ECG signal alone is not reliable enough to be used for R-peak derived cardiac vital signs however does not re-compute a more accurate value since no other

acceptable signal is available. With multiple signals, either another ECG lead or another signal entirely such as PPG, the proposed SQI can be extracted on each signal individually and then used to determine from which one does the most reliable cardiac vital sign come from and then output those values to the user. More sensors and more physiological information are an opportunity to output more reliable data and paint a fuller picture of the user-specific biological state.

* * *

Bibliography

[1] "IDC Reports Strong Growth in the Worldwide Wearables Market, Led by Holiday Shipments of Smartwatches, Wrist Bands, and Ear-Worn Devices." *International Data Corporation*. (2019). Web. <https://www.idc.com/getdoc.jsp?containerId=prUS44901819>.

[2] "Wearables Market to Be Worth $25 Billion by 2019." *CCS Insight*. (2019). Web. <https://www.ccsinsight.com/press/company-news/2332-wearables-market-to-be-worth-25-billion-by-2019-reveals-ccs-insight/>.

[3] "Global Wearable Technology Market Growth, Trends and Forecasts Through 2019-2024 - Smart Clothing to Hold Significant Share" *Business Wire*. (2019). Web. <https://www.businesswire.com/news/home/20190621005131/en/Global-Wearable-Technology-Market-Growth-Trends-Forecasts>.

[4] "Global Wearable Sensors Market Size: Industry Report, 2018-2025." *Grand View Research*. (2018). Web. <https://www.grandviewresearch.com/industry-analysis/global-wearable-sensor-marke>.

[5] "Global Wearable Device Market Size, Market Share, Application Analysis, Regional Outlook, Growth Trends, Key Players, Competitive Strategies and Forecast, 2018-2026." *Research and Markets*. (2018). Web. <https://www.researchandmarkets.com/research/3f8t2m/global_wearable?w=4>.

[6] Peake, J. M., G. Kerr, and J. P. Sullivan. "A Critical Review of Consumer Wearables, Mobile Applications, and Equipment for Providing Biofeedback, Monitoring Stress, and Sleep in Physically Active Populations." *Frontiers in Physiology* 9 (2018).

[7] "The Most Successful Wearables for Consumers." *Wearable Technologies*. (2016). Web. https://www.businesswire.com/news/home/20190621005131/en/Global-Wearable-Technology-Market-Growth-Trends-Forecasts.

[8] Vitaterna, M. H., J. S. Takahashi, and F. W. Turek. "Overview of circadian rhythms." *Alcohol Res Health* 25, no. 2 (2001): 85-93.

[9] Boudreau, P., G. Dumont, N. K. N. Y. Kin, C.-D. Walker, and D. B. Boivin. "Correlation of heart rate variability and circadian markers in humans." in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (2011): 681-682.

[10] Wyatt, J. K., A. R.-D. Cecco, C. A. Czeisler, and D.-J. Dijk. "Circadian temperature and melatonin rhythms, sleep, and neurobehavioral function in humans living on a 20-h day." *American Journal of Physiology, Regulatory and Integrative Physiology* 277, no. 4 (1999): R1152-R1163.

[11] Krauchi, K., and A. Wirz-Justice. "Circadian rhythm of heat production, heart rate, and skin and core temperature under unmasking conditions in men." *American Journal of Physiology-Regulatory and Integrative Physiology* 267, no. 3 (1999): R819-R829.

[12] Kalsbeek, A., C.-X. Yi, S. E. La Fleur, and E. Fliers. "The hypothalamic clock and its control of glucose homeostasis." *Trends in Endocrinology & Metabolism* 21, no. 7 (2010): 402-410.

[13] Ehlers, C. L., E. Frank, and D. J. Kupfer. "Social zeitgebers and biological rhythms: a unified approach to understanding the etiology of depression." *Archives of general psychiatry* 45, no. 10 (1988): 948-952.

[14] Kalsbeek, A., S. la Fleur, and E. Fliers. "Circadian control of glucose metabolism." *Molecular Metabolism* 3, no. 4 (2014): 372-383.

[15] Ehlers, C. L. "Social zeitgebers, biological rhythms and depression." *Clinical neuropharmacology* 15, no. Suppl 1 Pt A (1992): 44A-45A.

[16] Boudreau, P., W. H. Yeh, G. A. Dumont, and D. B. Boivin. "A Circadian Rhythm in Heart Rate Variability Contributes to the Increased Cardiac Sympathovagal Response to Awakening in the Morning." *Chronobiology International* 29, no. 6 (2012): 757-768.

[17] Sancar, A., L. A. Lindsey-Boltz, S. Gaddameedhi, C. P. Selby, R. Ye, Y. Y. Chiou, M. G. Kemp*, et al.* "Circadian clock, cancer, and chemotherapy." *Biochemistry* 54, no. 2 (2015): 110-23.

[18] Zhao, M., L. Zhang, Z. Wang, X. Wang, Y. Wang, H. Wei, R. Li, and Y. Du. "Dynamic analysis of blood pressure changes in progressive cerebral infarction." *International Health* 7, no. 4 (2015): 293-297.

[19] Takahashi, J. S., H. K. Hong, C. H. Ko, and E. L. McDearmon. "The Genetics of Mammalian Circadian Order and Disorder: Implications for Physiology and Disease." *Nature reviews. Genetics* 9, no. 10 (2008): 764-75.

[20] Alila07. "Pituitary Gland And Optic Chiasm." *Dreamstime*. Web. <www.dreamstime.com/royalty-free-stock-photos-pituitary-gland-optic-chiasm-image25994518>.

[21] Boivin, Diane. "Circadian Rhythms: What Are They?" *Centre for Study and Treatment of Circadian Rhythms, Douglas Research Centre*. (2013). Web. <www.douglas.qc.ca/info/circadian-rhythms-what-are-they>.

[22] Kräuchi, K., C. Cajochen, E. Werth, and A. Wirz-Justice. "Functional Link between Distal Vasodilation and Sleep-Onset Latency?". *American journal of physiology. Regulatory, integrative and comparative physiology* 278, no. 3 (2000): 741-8.

[23] Task Force of the European Society of, Cardiology, Pacing the North American Society of, and Electrophysiology. "Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use." *Circulation* 93, no. 5 (1996): 1043-65.

[24] Cohen, M. C., K. M. Rohtla, C. E. Lavery, J. E. Muller, and M. A. Mittleman. "Meta-Analysis of the Morning Excess of Acute Myocardial Infarction and Sudden Cardiac Death." *The American journal of cardiology* 79, no. 11 (1997): 1512-6.

[25] Kozak, M., L. KrIvan, and B. Semrad. "Circadian Variations in the Occurrence of Ventricular Tachyarrhythmias in Patients with Implantable Cardioverter Defibrillators." *Pacing and Clinical Electrophysiology* 26, no. 3 (2003): 731-35.

[26] Teng, X. F., and Y. T. Zhang. "The Effect of Applied Sensor Contact Force on Pulse Transit Time." *Physiological measurement* 27, no. 8 (2006): 675-84.

[27] Mukkamala, Ramakrishna, Jin-Oh Hahn, Omer T. Inan, Lalit K. Mestha, Chang-Sei Kim, Hakan Toreyin, and Survi Kyal. "Toward Ubiquitous Blood Pressure Monitoring Via Pulse Transit Time: Theory and Practice." *IEEE Transactions on Biomedical Engineering* 62, no. 8 (2015).

[28] Rakshit, Manas, and Susmita Das. "An Efficient Ecg Denoising Methodology Using Empirical Mode Decomposition and Adaptive Switching Mean Filter." *Biomedical Signal Processing and Control* 40 (2018): 140-48.

[29] Tobon V, Diana P., Tiago H. Falk, and Martin Maier. "Ms-Qi: A Modulation Spectrum-Based Ecg Quality Index for Telehealth Applications." *IEEE Transactions on Biomedical Engineering* 63, no. 8 (2016): 1613-22.

[30] Cuomo, S., G. De Pietro, R. Farina, A. Galletti, and G. Sannino. "A Revised Scheme for Real Time Ecg Signal Denoising Based on Recursive Filtering." *Biomedical Signal Processing and Control* 27 (2016): 134-44.

[31] Cheng, T. M., A. V. Savkin, B. G. Celler, S. W. Su, Wang Lu, Ieee Power, amp, and Mn Usa Energy Society General Meeting Minneapolis. "Nonlinear Modeling and Control of Human Heart Rate Response During Exercise with Various Work Load Intensities." *IEEE Transactions on Biomedical Engineering* 55, no. 11 (2008).

[32] Vandewalle, Gilles, Benita Middleton, Shantha M. W. Rajaratnam, Barbara M. Stone, Bjorg Thorleifsdottir, Josephine Arendt, and Derk-Jan Dijk. "Robust Circadian Rhythm in Heart Rate and Its Variability: Influence of Exogenous Melatonin and Photoperiod." *Journal of Sleep Research* 16, no. 2 (2007): 148-55.

[33] Krauchi, Kurt. "How Is the Circadian Rhythm of Core Body Temperature Regulated?". *Clinical Autonomic Research* 12, no. 3 (2002): 147-49.

[34] Agateller, Anthony Atkielski. "Schematic diagram of normal sinus rhythm for a human hearsy as seen on ECG (with English labels)." *SinusRhythmLabels*.svg. (2007). Web.

[35] "Cardiovascular Diseases (CVDs)." *World Health Organization*. (2017). Web. <https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)>.

[36] Gawlowska, J., J. K. Wranicz. "Norman J. "Jeff" Holter (1914-1983)." *Cardiology Journal* 16, no. 4 (2009): 1-2.

[37] Goldberg A. L., L. A. N. Amaral, L. Glass, J. M. Hausdorf, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals." *Circulation* 101, no. 23 (2003): e215-e220.

[38] Jeong Su, Lee, Heo Jeong, Lee Won Kyu, Lim Yong Gyu, Kim Youn Ho, and Park Kwang Suk. "Flexible Capacitive Electrodes for Minimizing Motion Artifacts in Ambulatory Electrocardiograms." *Sensors* 14, no. 8 (2014): 14732-43.

[39] Chi, Y. M., Jung Tzyy-Ping, and G. Cauwenberghs. "Dry-Contact and Noncontact Biopotential Electrodes: Methodological Review." *IEEE Reviews in Biomedical Engineering* 3 (2010).

[40] ANSI/AAMI/IEC 60601-2-47:2012/(R)2016. Medical electrical equipment – Part 2;47: Particular requirments for the basic safety and essential performance of ambulatory electrocadiographic systems.

[41] Pan, J., and W. J. Tompkins. "A Real-Time Qrs Detection Algorithm." *IEEE transactions on bio-medical engineering* 32, no. 3 (1985): 230-6.

[42] Hamilton, P. S., and W. J. Tompkins. "Quantitative Investigation of Qrs Detection Rules Using the Mit/Bih Arrhythmia Database." *IEEE transactions on bio-medical engineering* 33, no. 12 (1986): 1157-65.

[43] Zong, W., G. B. Moody, and D. Jiang. "A robust open-source algorithm to detect onset and duration of QRS complexes." In *Computers in Cardiology, 2003*, pp. 737-740. IEEE, 2003.

[44] Orphanidou, C., T. Bonnici, P. Charlton, D. Clifton, D. Vallance, and L. Tarassenko. "Signal-Quality Indices for the Electrocardiogram and Photoplethysmogram: Derivation and Applications to Wireless Monitoring." *IEEE Journal of Biomedical and Health Informatics* 19, no. 3 (2015).

[45] Orphanidou, C., and I. Drobnjak. "Quality Assessment of Ambulatory ECG Using Wavelet Entropy of the HRV Signal." *IEEE Journal of Biomedical and Health Informatics* 21, no. 5 (2017).

[46] Hayn, D., Jammerbund, B., & Schreier, G. (2012). QRS detection based ECG quality assessment. Physiological Measurement, 33, 1449–1461.

[47] Moody, B. E. "Rule-based methods for ECG quality control." In *2011 Computing in Cardiology*, pp. 361-363. IEEE, 2011.

[48] Allen, J., and A. Murray. "Assessing ECG signal quality on a coronary care unit." *Physiological measurement* 17, no. 4 (1996): 249.

[49] Langley, P., L. Y. Di Marco, S. King, D. Duncan, C. Di Maria, W. Duan, M. Bojarnejad, D. Zheng, J. Allen, and A. Murray. "An algorithm for assessment of quality of ECGs acquired via mobile telephones." In *2011 Computing in Cardiology*, pp. 281-284. IEEE, 2011.

[50] Behar, J., J. Oster, Q. Li, and G. D. Clifford. "ECG signal quality during arrhythmia and its application to false alarm reduction." *IEEE transactions on biomedical engineering* 60, no. 6 (2013): 1660-1666.

[51] Li, Q., R. G. Mark, and G. D. Clifford. "Robust Heart Rate Estimation from Multiple Asynchronous Noisy Sources Using Signal Quality Indices and a Kalman Filter." *Physiological Measurement* 29, no. 1 (2008): 15-32.

[52] Clifford, G. D., J. Behar, Q. Li, and I. Rezek. "Signal Quality Indices and Data Fusion for Determining Clinical Acceptability of Electrocardiograms." *Physiological Measurement* 33, no. 9 (2012): 1419-33.

[53] "CARESCAPE Patient Monitoring Module." *GE Healthcare*. Specification sheet. (2009). Web. <http://www.usmed-equip.com/content/datasheet/GE%20pdm%20spec%20sheet.pdf>.

[54] Saeed, M., C. Lieu, G. Raber, and R. G. Mark. "MIMIC II: a massive temporal ICU patient database to support research in intelligent patient monitoring." In *Computers in cardiology*, pp. 641-644. IEEE, 2002.

[55] Kostoglou, K., K. P. Michmizos, P. Stathis, D. Sakas, K. S. Nikita, and G. D. Mitsis. "Classification and Prediction of Clinical Improvement in Deep Brain Stimulation from Intraoperative Microelectrode Recordings." *IEEE Transactions on Biomedical Engineering* 64, no. 5 (2017): 1123-30

[56] Rankawat, Shalini A., and Rahul Dubey. "Robust Heart Rate Estimation from Multimodal Physiological Signals Using Beat Signal Quality Index Based Majority Voting Fusion Method." *Biomedical Signal Processing and Control* 33 (2017): 201-12.

[57] *Apple*. Retrieved from: Bell, K. "Apple Watch wearers in U.K. could wait years for ECG feature." (2018). Web. <https://cultofmac.com/579594/apple-watch-ecg-uk/>.

Appendices

Appendix A

Raw Polar HR data collected over Recording Periods 1 and 2 are presented in **Figure A-1** and **Figure A-2**, respectively.

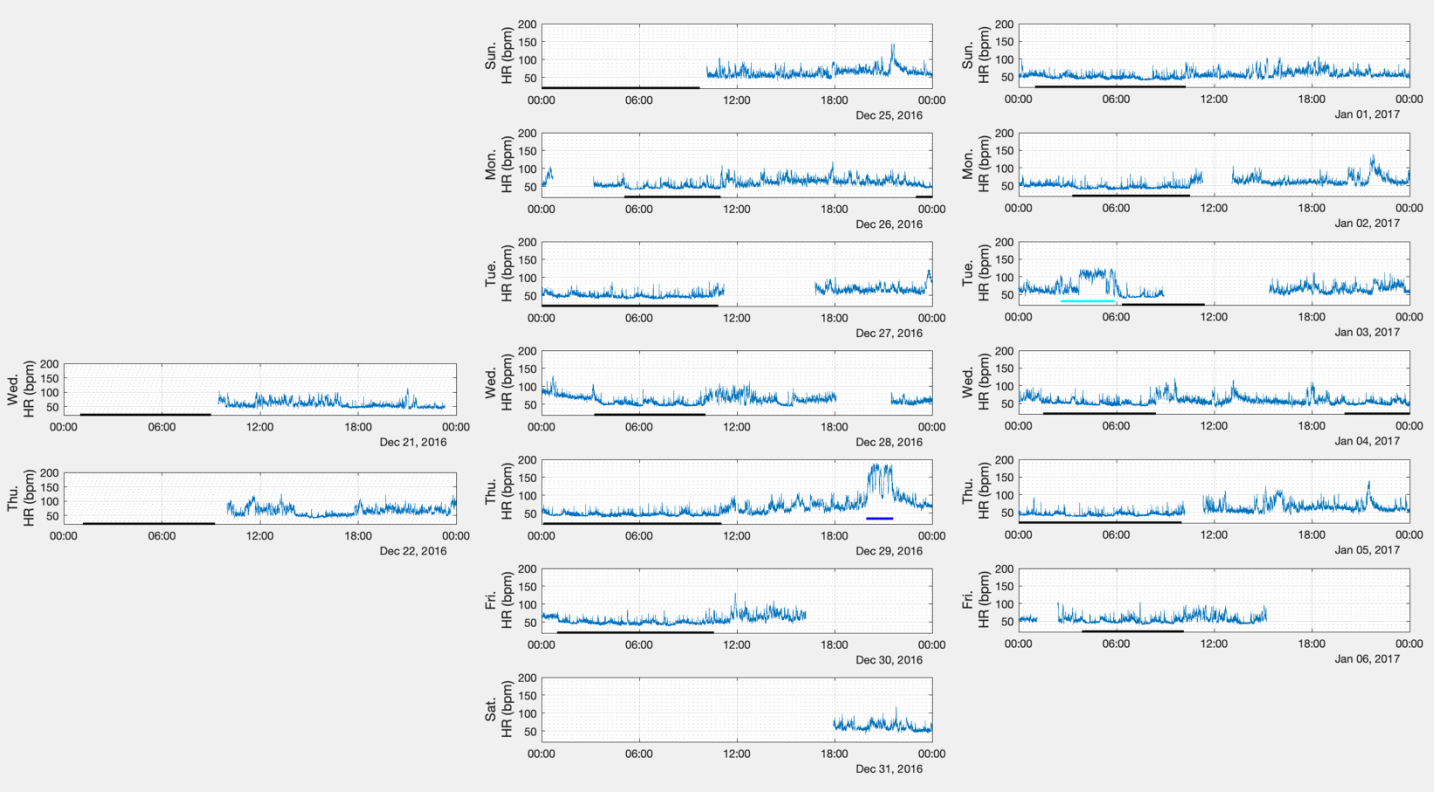

Figure A-1: Raw Polar HR data (cyan signal) collected over Recording Period 1 (15 days total with HR data), separated into days of the week and the 24-hour clock time. Solid bar lines represent intervals of external factors (sleep = black; exercise = blue; overnight + coffee = teal).

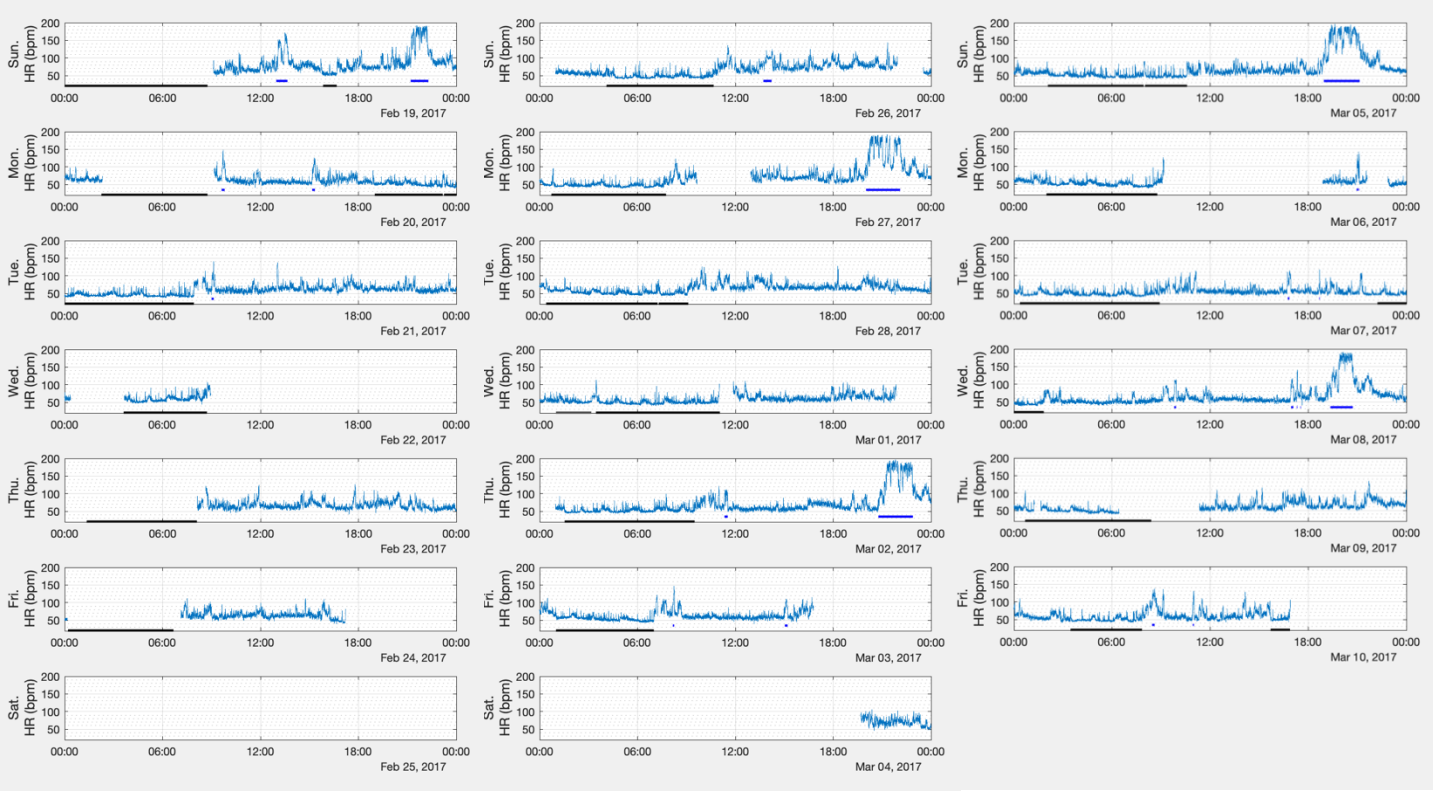

Figure A-2: Raw Polar HR data (cyan signal) collected over Recording Period 2 (19 days total with HR data), separated into days of the week and the 24-hour clock time. Solid bar lines represent intervals of external factors (sleep = black; exercise = blue).