



Universidad
Zaragoza

Ph.D. Thesis

**Robust Autonomic Nervous System
Monitorization Using Wearables
in Chronic Diseases**

Author

Diego Cajal Orleans

Supervisors

Raquel Bailón Luesma

Jesus Lázaro Plaza

Ph.D. in Biomedical Engineering

December, 2024

Robust Autonomic Nervous System Monitorization Using Wearables in Chronic Diseases

Diego Cajal Orleans, December 2024

Cover design: Iván Gil Martínez

This thesis has been developed in the frame of the following scientific projects: "Monitoring and assessment of obstructive sleep apnea and its outcomes through the processing of physiological signals of wearables" (PID2021-126734OB-C21); "Monitorización de la respiración mediante un prototipo de dispositivo "wearable" de tipo reloj de pulsera" (JIUZ2022-IAR-02); "Major Depressive Disorder monitoring system: development of a platform based on wearable devices in relevant scenarios" (PDC2022-133197-I00); "HEalthcare and MOnitoring DIgitalization in Chronic Illness management: application to multiple sclerosis and depression" (TED2021-131106B-I00); and "PPG-based respiratory monitoring" (Funding agency: Brainmatterz LLC at McKinney, TX, USA).

Contents

Abstract	I
Resumen y Conclusiones	II
Agradecimientos	IV
1 Introduction	1
1 Context and motivations	1
2 Objectives.....	3
3 Structure of the thesis	5
Bibliography	8
2 Wearables, biosignals and physiology	11
1 The wearable revolution.....	11
2 Biosignal acquisition and interpretation	13
2.1 Pulse photoplethysmography	14
2.2 Electrocardiography	16

2.3	Peripheral venous pressure	19
2.4	Auxiliar signals	20
3	Physiology	22
3.1	Autonomic nervous system	22
3.2	Autonomic control of the cardiorespiratory system	25
3.3	Heart rate variability	28
	Bibliography	33
3	Pathology under study	43
1	Chronic mental health conditions	43
1.1	Relationship with stress reactivity	45
1.2	Wearable monitoring	46
1.3	Main challenges	46
2	Obstructive sleep apnea	48
2.1	Wearable monitoring	51
2.2	Main challenges	52
	Bibliography	55
4	General methodology for robust ANS monitorization ..	65
1	Methodological framework	65
2	Artifact removal in PPG signals	69
2.1	Hjorth parameters	72
3	Signal quality assessment	75

4	Event detection	76
4.1	Event detection in ECG signals.....	76
4.2	Event detection in PPG signals.....	77
5	Event correction.....	79
5.1	Ectopic beats	81
6	Variability metrics.....	82
6.1	Time domain	82
6.2	Frequency domain (I). Spectral estimation	84
6.3	Frequency domain (II). Component separation	88
6.4	Frequency domain (III). Metrics computation.....	94
6.5	Lagged Poincaré plots	95
	Bibliography	96
5	Respiration-guided ANS evaluation from wrist PVP....	105
1	Objectives and motivations.....	105
2	Materials and methods.....	106
2.1	Signal delineation	106
2.2	Respiratory rate estimation	108
2.3	Respiration-guided parasympathetic assessment	109
3	Results.....	110
4	Discussion	111
4.1	Findings.....	111
4.2	Limitations	112

5	Conclusion.....	113
	Bibliography	113
6	Effects of missing data in HRV metrics	117
1	Objectives and motivations	117
2	Introduction	118
2.1	Related work	119
2.2	Objectives and motivations	120
3	Materials and Methods.....	121
3.1	Simulation of missing beats.....	121
3.2	Apple Watch dataset.....	124
3.3	Missing data detection.....	126
3.4	Correction methods	126
3.5	HRV metrics.....	129
3.6	Statistical analysis	130
4	Results.....	131
4.1	Time-domain metrics	131
4.2	Frequency-domain metrics computed via Fourier transform.....	137
4.3	Frequency-domain metrics computed via Lomb-Scargle periodograms	141
4.4	Poincaré Plots	144

5	Discussion	149
5.1	Time-domain metrics	150
5.2	Frequency-domain metrics computed via Fourier transform.....	151
5.3	Frequency-domain metrics computed via Lomb-Scargle periodograms	151
5.4	Poincaré Plots	154
5.5	Iterative gap-filling approach	154
5.6	Limitations	155
6	Conclusions.....	156
	Bibliography	157
7	Mental health monitoring using wearables	163
1	Objectives and motivations	163
2	Experiment	164
2.1	Stress protocol.....	166
3	Methods	168
3.1	Filtering and artifact removal	168
3.2	Pulse detection	168
3.3	PRV metrics	169
3.4	Statistical analysis	170
4	Results	170
4.1	Agreement	170

4.2	Stress reactivity	176
5	Discussion	177
5.1	SCPPG for PRV monitoring	177
5.2	Stress reactivity	180
5.3	Limitations	182
6	Conclusions	183
	Bibliography	183
8	Early detection of sleep apnea (I)	191
1	Objectives and motivations	191
2	Introduction	192
3	Methods	193
3.1	Dataset	193
3.2	Signals for segment classification	193
3.3	References for performance evaluation	194
3.4	Segment classification models	198
3.5	OSA stratification by CVHRI	199
3.6	Performance analysis	200
4	Results	202
4.1	Segment classification results	202
4.2	CVHRI correlation with AHI	203
4.3	OSA stratification by CVHRI	203

5	Discussion	207
5.1	Segment classification	207
5.2	CVHRI correlation with AHI	210
5.3	OSA stratification by CVHRI	210
5.4	Limitations	212
6	Frequency-domain PRV metrics as predictors	213
7	Conclusion	215
	Bibliography	215
9	Early detection of sleep apnea (II)	221
1	Methods	221
1.1	Dataset	221
1.2	Peripheral oxygen saturation estimation	223
2	Results	225
3	Discussion	227
4	Conclusions	229
	Bibliography	229
10	Integrating findings: from theory to practice	233
1	Primary findings	233
2	Limitations	235
3	Future lines	236
4	Smarthone app	237

5	Conclusions.....	242
	Bibliography	244
	List of Figures.....	256
	List of Tables.....	260
	List of Acronyms.....	260
	List of Publications	263

Abstract

The appearance of wearable devices in the field of biomedicine marks an important milestone in the evolution of technology applied to health care and personal well-being. These devices have revolutionized the way the physiological parameters and daily activities of the individual are monitored. Their development is based on the advanced miniaturization of sensors and signal processing systems, allowing real-time capture and analysis of critical data such as heart rate, blood oxygenation, and physical activity level, among others. This innovation not only facilitates constant and preventive health monitoring, but also opens new avenues for the personalization of medical treatment and rehabilitation.

Wearable devices come with certain limitations that must be carefully considered in the design of signal processing algorithms. One of the main challenges is the quality of the data collected in uncontrolled environments. Unsupervised recordings can be susceptible to variations in device placement, extrinsic movements, and varying environmental conditions, which can result in inaccurate or incomplete measurements, especially when these recordings are made during daily life. Additionally, the very nature of the sensors used in these devices, as well as the biological signals they record, tends to produce data with a higher level of noise and artifacts, since there is a trade-off situation between the optimal place in signal quality and comfortability terms. This is particularly noticeable in signals such as photoplethysmography.

To counteract these drawbacks, it is imperative to adapt and develop robust analysis methods that can handle these inherent imperfections in the data.

This includes preprocessing techniques for artifact removal, as well as more sophisticated analysis algorithms that can extract useful information even from incomplete data. This thesis presents several methods and their applications in the context of chronic disease monitoring using wearable devices. Specifically, it investigates applications for identifying sleep apnea and monitoring depression and anxiety.

Resumen y Conclusiones

La aparición de los dispositivos wearables en el ámbito de la biomedicina marca un hito importante en la evolución de la tecnología aplicada al cuidado de la salud y el bienestar personal. Estos dispositivos han revolucionado la forma en que se monitorizan los parámetros fisiológicos y las actividades cotidianas del individuo. Su desarrollo se basa en la miniaturización avanzada de sensores y sistemas de procesamiento de señales, permitiendo la captura y análisis en tiempo real de datos críticos como la frecuencia cardíaca, la oxigenación sanguínea y el nivel de actividad física, entre otros. Esta innovación no solo facilita una vigilancia constante y preventiva de la salud, sino que también abre nuevas vías para la personalización del tratamiento médico y la rehabilitación.

Los dispositivos wearables conllevan ciertas limitaciones que deben ser cuidadosamente consideradas en el diseño de algoritmos de procesamiento de señales. Uno de los desafíos principales es la calidad de los datos recogidos en entornos no controlados. Las grabaciones no supervisadas, realizadas en la vida cotidiana del usuario, pueden ser susceptibles a variaciones en la colocación del dispositivo, movimientos extrínsecos, y condiciones ambientales variables, lo cual puede resultar en mediciones imprecisas o incompletas. Adicionalmente, la naturaleza misma de los sensores utilizados en estos dispositivos, así como de las señales biológicas que registran, tiende a producir datos con un mayor nivel de ruido y artefactos. Esto es particularmente notable en señales como la fotopletimografía.

Para contrarrestar estos inconvenientes, es imperativo adaptar y desarrollar métodos de análisis robustos que puedan manejar estas imperfecciones

inherentes a los datos. Esto incluye técnicas de preprocesamiento para la eliminación de artefactos, así como algoritmos de análisis más sofisticados que puedan extraer información útil incluso de datos incompletos. En esta tesis se presentan algunos de estos métodos, así como aplicaciones de los mismos en la monitorización de enfermedades crónicas con dispositivos wearables. En concreto, se presentan aplicaciones para el screening de apnea del sueño y para la monitorización de la depresión y la ansiedad.

Agradecimientos

Esta tesis no sería posible sin un gran número de personas. Espero no dejarme a nadie.

Quiero empezar agradeciendo el esfuerzo de mis directores, coautores de todos los trabajos que forman este manuscrito. Raquel, has estado presente en todas mis fases formativas en esta Universidad, desde primero de carrera, pasando por el máster y el TFM hasta este momento. Eres incansable, dedicada y aprendo de cada conversación contigo. Espero seguir haciéndolo. No se me borra de la cabeza el día que me acerqué a tu despacho a preguntarte por una oferta de TFM y salí con un plan de doctorado. No me arrepiento. De Jesús también puedo decir que he aprendido un gran número de cosas. Aunque todavía recuerdo con terror su corrección de mi primer artículo -estaba todo en rojo-, forma parte inseparable de mi crecimiento profesional en estos años. Con él compartí el que de momento es de largo el mejor congreso en el que he estado. En este espacio también quiero agradecer a Edu, que si bien no ha sido un director más ha estado cerca; a David, especialmente presente en mis primeros pasos; y a Pablo, cuyas indicaciones encierran siempre una gran sabiduría.

No puedo menos que dar las gracias a todos mis compañeros del día a día. Especialmente a los que han estado durante todo este viaje. Cristina y Pablo, habéis sido los mejores compañeros que uno puede desear. Vuestras conversaciones han sido imprescindibles para mantener el ánimo, y de ellas han surgido también nuevas ideas y enfoques que hoy se plasman en estas líneas. A este núcleo también pertenece Rodrigo, que aunque no ha estado desde el primer día ha sido un compañero de fatigas en la etapa

intermedia-final. No me olvido tampoco del resto de mis compañeros de laboratorio, que por desgracia ya no están por aquí. Spyros, Javi, Carmen, Estela, Andrés... os tengo mucho cariño.

En general quiero agradecer estos años a todo el grupo BSICoS. Habéis sido unos grandes compañeros de los que he aprendido innumerables cosas durante estos años. Las tardes de deporte han sido un desahogo inigualable de la presión del doctorado.

También quiero agradecer a la gente del STAR group de Kuopio, en Finlandia. Sigue indeleble el recuerdo de aquella comida con Samu, donde todas las cabezas se giraron hacia mí cuando se propuso que nuestra colaboración incluyese estancias de investigación. Agradezco mucho la confianza que pusieron en mí para esta gran -y fría- experiencia. Gracias a Salla, Marika, Juuso y Tim por hacer mi estancia más amena y permitirme vivir la experiencia finlandesa completa.

Me reservo las últimas líneas para aquellos que más quiero. A mis padres, que me han querido y apoyado durante tantos años. Siempre me pusieron todas las facilidades para llevar mis estudios hasta el final, que no se limitaron a la universidad. Gracias a ellos soy lo que soy. Carlos, no te pongas celoso, estas palabras también van por ti. En estos momentos recuerdo con especial cariño a los que no están, mis abuelos, que seguro que estarían orgullosísimos de esto, como lo estuvieron de todo lo que hacía su nieto. También te lo agradezco a ti, Paula, por aguantar todos estos años a mi lado a pesar de ser un cabezón. Han sido años de grandes cambios y grandes decisiones, que me han hecho crecer como persona a tu lado. Espero que sean muchos años más. Acabo estas líneas agradeciendo a mis amigos -no pongo nombres porque por suerte para mí sois muchos-, por todos esos momentos de esparcimiento, confidencias y "comentadas" en general. Lleváis conmigo toda la vida. Os quiero mucho.

1 Context and motivations

Wearable devices have become key tools in monitoring health and wellness. Their application spans various areas, from preventive medicine to chronic disease management, promoting greater personalization and precision in healthcare. Wearable devices, which embody a spectrum of sensors and computational capabilities, have transitioned from niche applications to mainstream adoption due to their potential to offer continuous, real-time insights into an individual's physiological states and activity levels. Continuous monitoring of physiological parameters such as heart rate, blood pressure, glucose levels, electrodermal activity, and blood oxygen saturation allows early detection of anomalies, facilitating timely medical interventions. For example, wearables can detect irregular heartbeats or poor sleep, having the potential to alert both the user and healthcare providers to potential issues before they become critical. Wearables also track physical activity, sleep patterns, and other lifestyle factors. By providing real-time feedback, they encourage healthier habits and increased physical activity. Features such as step counting, calorie tracking, and sleep analysis help individuals set and achieve personal health goals [1, 2]. In addition, some wearables offer stress monitoring and guided breathing exercises, which contribute to mental well-being. In general, the integration of wearables in health and

wellness represents a significant shift towards proactive and personalized healthcare, empowering individuals to take an active role in managing their health.

The data generated by wearable devices are invaluable for medical research, offering large longitudinal datasets that were previously unattainable. This has implications for personalized medicine, where treatments can be tailored to the individual's unique physiological makeup and lifestyle, potentially improving outcomes and reducing healthcare costs. In addition, the comfort and convenience of noninvasive techniques suggest widespread adoption by the general population, facilitating early screening and detection. Central to this evolution is the integration of biomedical engineering with signal processing techniques, which transform raw data into actionable health metrics. This convergence facilitates a more nuanced understanding of the human body, which extends beyond traditional clinical settings to everyday life. However, this rapid advancement is highly challenging. The issues related to data privacy, security, and accuracy of wearable devices are areas of ongoing concern. The precision of data interpretation, crucial for making health-related decisions, is heavily based on the sophistication of the algorithms used, which must be continually refined in line with advances in our understanding of human physiology and the environment. This thesis focuses on developing robust methodologies to estimate informative health parameters from wearable device signals.

The continuous monitoring of biomedical signals is both the primary advantage and the main challenge faced by wearable devices. Movements, variations in environmental conditions, and daily physiological changes generate numerous scenarios that must be considered. Motion artifacts and low signal-to-noise ratios are frequent issues, and current algorithms are not entirely reliable in addressing these challenges. Many of the commonly used metrics were developed prior to the advent of wearable devices and were based on signals obtained from clinical devices in controlled environments. Consequently, these initial algorithms assumed a certain level of signal quality, with significantly fewer artifacts present. Therefore, signal processing methodologies have had to evolve to account for these

new conditions. This work outlines the primary challenges that must be addressed in the short term throughout the entire signal processing chain and proposes solutions for some of these critical points.

This proposal extends beyond the methodology to include practical applications. In addition to examining the algorithms themselves, their application in two promising areas for wearables is addressed: mental illnesses, such as depression and anxiety; and obstructive sleep apnea. Mental illnesses have become a focal point in contemporary society due to their high incidence and severity. Disorders such as major depression and generalized anxiety are highly disabling. The implementation of wearables as a monitoring tool is presented as a promising approach to personalize and track treatment. In addition, the fields of psychiatry and psychology may benefit from the numerical metrics derived from biosignals, which would supplement their primarily subjective tools and measures. On the other hand, obstructive sleep apnea, a disease with a very high incidence but significant underdiagnosis, is particularly highlighted. Wearables possess the ability to analyze biosignals during the night, including those closely related to sleep apnea, such as peripheral oxygen saturation. This thesis analyzes their use as a screening tool and proposes a new and promising index to alert users.

2 Objectives

The objective of this thesis is to investigate robust and noninvasive methods for assessing the autonomic nervous system using wearable devices. This study draws on all previous research derived from electrocardiograms, as well as more recent pulse rate variability studies based on photoplethysmography. The signals acquired from the wearable devices have distinctive characteristics that must be addressed specifically to achieve comparable results. Progress in these objectives brings society closer to a deeper understanding of the human body, particularly the autonomic nervous and cardiovascular systems, through extensive biosignal recordings from both patient and general populations. In addition, the applications of wearable

devices are being investigated for the screening of obstructive sleep apnea and monitoring stress reactivity, one of the physiological mechanisms altered in mental illnesses such as major depression and generalized anxiety, proposing their use in the follow-up of these diseases.

The specific objectives of this thesis are:

1. *General methods for reliable ANS assessment using wearable devices*

Long-term unsupervised recordings (ideally 24 hours a day, 7 days a week) exposed to motion and changing conditions present a specific challenge for wearable devices. However, some of these issues are common to most of them, so it is possible to outline a common methodology. The primary objective of the thesis is to explore the impacts of missing data in order to ensure reliable monitoring.

2. *Smartphone app development for PPG recording*

The flashlight and camera of a smartphone can be employed to obtain the pulse photoplethysmography signal. This technology allows for the use of this device for monitoring some health-related parameters. In contrast to other wearables, such as smartwatches, it does not allow continuous recording while performing daily tasks. Additionally, it has specific challenges that are detailed in this thesis, such as a larger amount of motion artifacts. However, the universality of smartphones is an advantage that makes it a very attractive technology.

3. *Application of wearables to mental illness*

Mood disorders are one of the major health problems in today's society. The introduction of wearables for monitoring these patients can provide an additional layer of information for professionals, adding objective data, such as stress reactivity assessment measures, to a field where analysis is almost exclusively based on subjective parameters. The objective is to analyze whether stress reactivity can be evaluated with a smartphone in the same way as it is done with commercial pulsioximeter devices.

4. *Application of wearables to obstructive sleep apnea screening*

Obstructive sleep apnea is highly undiagnosed and its gold standard for diagnosis is polysomnography, which results in a cumbersome procedure. During this test, the patient has to spend one night in a hospital monitored by a variety of sensors. The path to this test involves self-detection of symptoms, such as daytime sleepiness, which, when reported to the physician, can lead to a recommendation to perform the test. Therefore, automatic alert and screening systems may greatly enhance the identification of a disease that poses significant long-term risks. This thesis explores the use of signals from commercial wearable devices, such as pulse photoplethysmography and peripheral oxygen saturation, as inputs for a sleep apnea screening model.

3 Structure of the thesis

The thesis is organized as follows:

- Chapter 2 presents an introduction to wearable devices in the context of health monitoring, along with the biosignals they provide and related physiology. This chapter aims to provide the reader with the tools to understand what information these devices provide and how they can be used to assess the autonomic nervous system through heart rate variability.
- Chapter 3 focuses on the examination of the pathologies under study: mental illness and obstructive sleep apnea. Although the core of the thesis emphasizes general methodologies relevant to various applications, practical instances are included to demonstrate how robust analysis of biosignals from wearable devices can enhance conventional techniques.
- Chapter 4 is dedicated to the methods shared by the following chapters, with special emphasis on the description of the processing chain as a whole, composed of distinguishable but interdependent parts, from signal acquisition to the computation of heart rate variability metrics.

- Chapter 5 is a small study conducted with an innovative device: a band with a piezoelectric sensor that records peripheral venous pressure. Since respiratory movements significantly influence venous return, metrics derived from this signal have been utilized to evaluate the parasympathetic autonomic system. The research described in this chapter generated the following publication:
 - **Diego Cajal**, David Hernando, Jesús Lázaro, Eduardo Gil, Annie Alvis, Monica Polcz, Bret Alvis, Kyle Hocking, Colleen Brophy & Raquel Bailón. *Parasympathetic Characterization Guided by Respiration From Wrist Peripheral Venous Pressure Waveform*. 2020 Computing in Cardiology, Rimini, Italy, 2020, pp. 1-4. **Oral communication.**
- Chapter 6 focuses on exploring the impact of data loss on heart rate variability metrics. The study includes a simulation to assess the degradation of various metrics in both the time and frequency domains, together with Poincaré plots. In addition, a gap-filling technique has been introduced to address data loss and ensure reliable metrics. This technique has been compared with other methods through simulation and a real-world application using an Apple Watch. The research described in this chapter generated the following publication:
 - **Diego Cajal**, David Hernando, Jesús Lázaro, Pablo Laguna, Eduardo Gil & Raquel Bailón. *Effects of Missing Data on Heart Rate Variability Metrics*. Sensors 2022, vol. 22, no. 15, 5774. **Journal article. Corresponding author.**
- Chapter 7 investigates the use of pulse photoplethysmography obtained using a smartphone camera to monitor patients with mental illness such as depression and anxiety. The application has been developed specifically for this work, applying the most robust error-correction methods from the previous chapter. This work currently continues with the development of a system that combines the mobile

application with cloud-based metric computation and a user interface for healthcare professionals where results are displayed. The aim of this project is to provide numerical and objective metrics in a field of medicine primarily guided by subjective evaluations. The research described in this chapter generated the following publications:

- **Diego Cajal**, Concepción de la Cámara, Mar Posadas-de Miguel, Noel Torrijos, Óscar Nadal, Teresa Blanco, Sara Siddi, Pablo Armañac, Eduardo Gil, Jesús Lázaro & Raquel Bailón. *Evaluation of Stress Response Using Smartphone PPG for Anxiety and Depression Monitoring* **Journal article. Corresponding author. Pending revision.**
- **Diego Cajal**, Mar Posadas-de Miguel, Concepción de la Cámara, Spyridon Kontaxis, Jesús Lázaro & Raquel Bailón. *Smartphone PPG Validation for a Depression Assessment Protocol*. 2022 IEEE e-Health and Bioengineering Conference (EHB), Iasi, Romania, 2022, pp. 1-4. **Oral communication.**
- **Diego Cajal**, Mar Posadas-de Miguel, Concepción de la Cámara, Spyridon Kontaxis, Jesús Lázaro & Raquel Bailón. *Smartphone PPG Validation for a Depression Assessment Protocol*. I Young Researchers Meeting CIBERESP-BBN 21-22 November 2022. **Poster.**
- Chapters 8 and 9 are dedicated to the study of a screening tool for obstructive sleep apnea using a pulse photoplethysmographic signal and peripheral oxygen saturation, common signals from wearable devices. Sleep apnea is widespread and significantly underdiagnosed, which can cause serious complications. Creating this kind of application can be beneficial in preventing complications and improving patient well-being. The research described in this chapter generated the following publications:
 - **Diego Cajal**, Eduardo Gil, Pablo Laguna, Carolina Varon, Dries Testelmans, Bertien Buyse, Chris Jensen, Rohan Hoare, Raquel

Bailón & Jesús Lázaro. *Obstructive Sleep Apnea Screening by Joint Saturation Signal Analysis and PPG-derived Pulse Rate Oscillations*. IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 1, pp. 228-238, 2024. **Journal article. Corresponding author.**

- **Diego Cajal**, Eduardo Gil, Pablo Laguna, Carolina Varon, Dries Testelmans, Bertien Buyse, Chris Jensen, Rohan Hoare, Raquel Bailón & Jesús Lázaro. *Sleep apnea severity stratification by an FFT-based PPG-derived index*. 2022 12th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO), Vysoké Tatry, Štrbské Pleso, Slovakia, 2022, pp. 1-2. **Oral communication.**
- **Diego Cajal**, Eduardo Gil, Pablo Laguna, Carolina Varon, Dries Testelmans, Bertien Buyse, Chris Jensen, Rohan Hoare, Raquel Bailón & Jesús Lázaro. *Estratificación de la severidad de la apnea del sueño mediante un índice derivado de PPG basado en FFT*. Jornada de Jóvenes Investigadores del I3A, vol. 10 (Actas de la XI Jornada de Jóvenes Investigadores del I3A – 16 de junio de 2022) **Oral communication.**
- Chapter 10 is dedicated to summarizing the conclusions and main findings of the thesis. It recapitulates the key objectives of the research and synthesizes the most significant discoveries from each chapter. It also reflects on the limitations of the study, proposes directions for future research, and concludes by emphasizing the importance of developing objective noninvasive tools for health diagnosis and monitoring in the context of personalized healthcare and emerging medical technologies.

Bibliography

- [1] Y. Jain, D. Chowdhury, and M. Chattopadhyay. Machine learning based fitness tracker platform using MEMS accelerometer. In *2017 International Conference on Computer, Electrical and Communication Engineering, ICCECE 2017*, 2017.
- [2] H.J. Tricás-Vidal, M.O. Lucha-López, C. Hidalgo-García, M.C. Vidal-Peracho, S. Monti-Ballano, and J.M. Tricás-Moreno. Health Habits and Wearable Activity Tracker Devices: Analytical Cross-Sectional Study. *Sensors*, 22(8), 2022.

Wearables, biosignals and physiology

1 The wearable revolution

Years following the creation of the first wireless electrocardiogram by Polar Electro in 1977, the entry of smartphone manufacturers in the mid-2010s greatly contributed to the expansion and advancement of wearable devices [1]. Today, the prominent brands dominating this market segment include giants such as Apple, Google (Fitbit), Garmin, Polar, Samsung, and Xiaomi. Pioneering devices such as the Garmin 101 have paved the way for the development of progressively advanced models, which now incorporate an extensive array of sensors, facilitating a broader spectrum of applications. Contemporary devices are often equipped with optical sensors capable of recording pulse photoplethysmography and monitoring oxygen saturation levels. Many of them also integrate electrical sensors designed for electrocardiography and electrodermal activity measurement. This suite of sensors is further enhanced by the incorporation of temperature sensors, pedometers, altimeters, gyroscopes, and GPS modules. Together, these sensors allow the tracking of various health and fitness metrics, such as heart rate, stress levels, sleep patterns, steps taken, distance traveled, speed, and elevation gain [2].

Health and fitness remain the main applications of wearable technology [3]: features such as fall detection on smart watches enable prompt identification

and notification of caregivers, addressing a major health concern with serious consequences, particularly for the elderly [4]; consistent monitoring of physical activity and calorie consumption can encourage individuals to adjust their exercise and diet routines, thus fostering a healthy lifestyle [5]; observing physical activity and vital signs has demonstrated advantages for patients with chronic diseases [6, 7]; etc. Monitoring applications have transformed the healthcare sector from reactive to proactive prevention-oriented interventions, allowing users to have greater responsibility and control over their health [8].

Several categories of devices exist: smartwatches, fitness bands, fitness trackers, smart shoe insoles, wearable biometric tracking devices, and smart garments. Within the purview of this thesis, smartphones are classified as wearable devices. These instruments, which generally serve as the central nodes for connectivity with additional gadgets, are proficient in executing accelerometry, positioning, and gyroscopic measurements. Moreover, smartphones possess the capability to capture pulse photoplethysmographic signals through the camera and conduct voice characteristic analysis via the microphone. Furthermore, fingerprint readers enable secure authentication for access or signal recording, thus ensuring the protection of privacy. The ubiquitous presence of smartphones presents significant potential for this technology [9].

A recent market analysis suggested that the demand for wearable devices would grow if their practicality and perceived ease of use were improved [8]. Although designing wearable devices is outside the scope of this thesis, researchers should consider these factors when creating new metrics or adapting existing ones for wearables. Wearable devices should increasingly incorporate novel health and wellness monitoring metrics that are comprehensible and straightforward to interpret for end users. This strategy could foster a positive feedback loop, where wider use of wearables produces more data, thus facilitating the investigation of new metrics or the addition of existing ones in innovative contexts.

2 Biosignal acquisition and interpretation

The utility of wearable devices is determined by their ability to record various biological signals, which are continuously generated by the human body during normal functioning. These signals can originate from diverse sources such as electrical, chemical, acoustic, and mechanical [10]. Although invasive wearables, such as subcutaneously implantable chips, may represent promising future developments, the focus on noninvasive techniques reflects a pragmatic approach given their current predominance and ease of adoption by the general population. The acquisition, processing, and analysis of signals obtained through noninvasive methods, particularly for the monitoring of chronic diseases, are the focus of this thesis. Specifically, attention is given to three types of signals that can be recorded on the body's surface: pulse photoplethysmography, electrocardiography, and peripheral venous pressure.

Medical specialists are trained in visually evaluating certain biosignals, such as electrocardiograms, to identify conditions such as arrhythmias. However, signal processing methods are essential for revealing concealed details and even during visual analysis, eliminating noise and correcting the baseline is vital. The varied nature of biosignals allows for their recording using optical (pulse photoplethysmography, peripheral oxygen saturation), electrical (electrocardiography, electromyography, electroencephalography, electrodermal response), mechanical (plethysmography, airflow measurement, ballistocardiography), acoustic (phonocardiogram, voice recording), and chemical sensors (glucose measurement). Often, a single device can record multiple signals simultaneously, enabling the integration of various sources of information into a cohesive model. This diversity necessitates a wide array of techniques for researchers, from preprocessing tasks like filtering and normalization to integrating data into complex models, including deep learning algorithms. Intermediate processes pose significant challenges, such as pulse detection in photoplethysmographic and electrocardiographic signals, managing data loss, and separating components linked to different signal sources. Before detailing the methodology used in this thesis, which

is presented in Chapter 4, it is crucial to describe the signals to be analyzed. A thorough understanding of the nature, origin, and physiological relationships of these biosignals is essential to uncover the underlying information. Likewise, insights gained from biosignal analysis must contribute to a deeper understanding of the natural processes of the human body that produce these signals.

2.1 Pulse photoplethysmography

Pulse photoplethysmography (PPG) represents an essential biosignal within the context of wearable technology, being widely incorporated in devices such as smartwatches, fitness trackers, and armbands (see Figure 1). PPG is a noninvasive technique employed to measure changes in blood volume within the microvasculature of tissues. The method operates by illuminating the skin with a light source and detecting variations in the intensity of the reflected or transmitted light. These variations correspond to changes in blood volume that occur during the cardiac cycle.

The PPG sensor comprises two essential components: a light source and a photodetector. The light source emits light directed toward the skin, while the photodetector measures the amount of light that is either reflected or transmitted (see Figure 2). Hemoglobin present in blood exhibits light absorption properties, showing particular sensitivity to wavelengths within the near-infrared and visible spectrum. Throughout the cardiac cycle, pulsatile changes in blood volume within the skin's blood vessels lead to variations in the local hemoglobin concentration, which in turn affect how light is absorbed. During systole, an increase in blood volume within the tissues results in heightened light absorption, leading to a decrease in the intensity of the detected light. Conversely, during diastole, the blood volume decreases, allowing more light to be detected.

The PPG signal provides insights into variations in blood flow and the oxygenation of peripheral tissues. The signal comprises a pulsatile component, which correlates with heartbeats, and a continuous component, which is associated with light absorption by tissues and the non-pulsatile



Figure 1: Examples of wearable devices that use PPG sensors. *From top left, clockwise: Apple Watch PPG sensor using green LEDs; Polar OH1+ armband; a smartwatch sharing pulse rate estimations with a smartphone; and a runner using a fitness tracker.* From [11–14].

blood component. PPG can provide information on several cardiovascular parameters, including heart rate, heart rate variability, oxygen saturation, and arterial stiffness. Peripheral oxygen saturation (SpO_2) is determined by pulse oximetry, a method that employs two light wavelengths, one red and one infrared, to differentiate between the light absorption properties of oxygenated and deoxygenated hemoglobin. The ratio of the light intensities detected at these two wavelengths is utilized to compute SpO_2 . More details of this process are given in Chapter 9.

Several factors can impact the accuracy of these measurements, including patient movement, inadequate peripheral perfusion, dark nail polish, and intense ambient light conditions. Smartphones can also measure PPG by using the camera and flash. When the fingertip is placed over the camera and flash, the device captures light variations by averaging the intensity of each frame. Health applications then process these signals to calculate heart rate

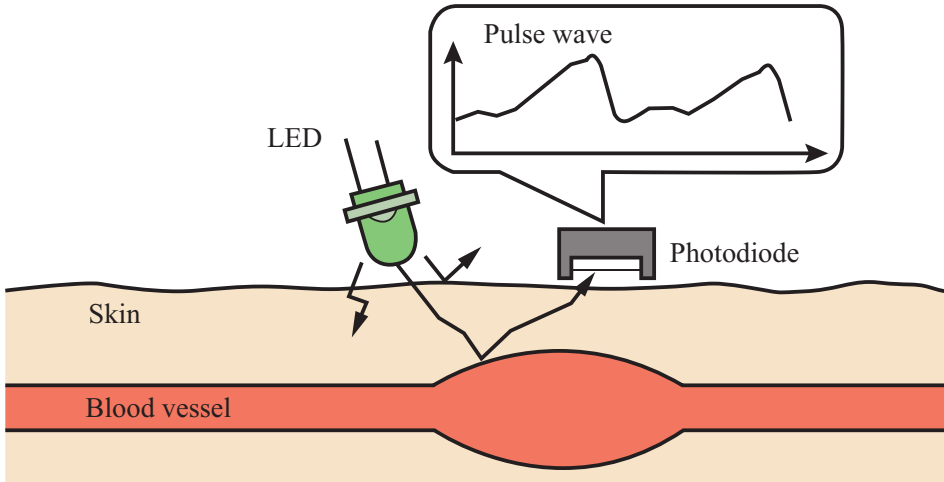


Figure 2: Operation of a reflection PPG. This is the type of PPG provided by smartwatches, smartphones and devices placed on the forehead, arm and temple. It is opposed to the transmittance PPG, where the photodiode is facing the LED, provided by fingertip pulse oximeters and earlobe devices. Adapted from [15].

and, in certain cases, heart rate variability. The accuracy of these smartphone measurements may be compromised by the same factors that affect pulse oximeters or smartwatch readings, with additional complications arising from movement and applied pressure. Moreover, current smartphone cameras lack infrared capabilities, preventing direct SpO_2 measurement, although attempts have been made to estimate it from PPG signals [16]. Despite these challenges, the ease of use and accessibility of smartphones make them a powerful tool for basic health monitoring.

2.2 Electrocardiography

Cardiac myocytes are activated by electrical impulses that trigger muscle contraction and act as transmitters in the nervous system. These impulses create an action potential through muscle tissue, facilitated by ion exchanges. Each phase of the cardiac cycle corresponds to a specific waveform in the ECG (see Figure 3). Atrial systole at the beginning of the heartbeat generates the P wave, followed by a brief pause that corresponds to the

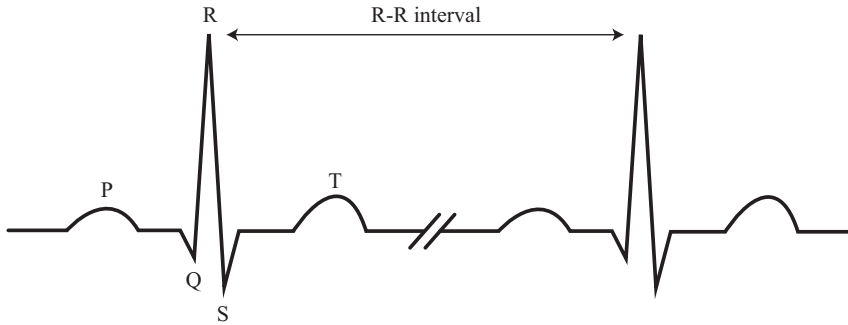


Figure 3: ECG waveform. The P wave corresponds with atrial depolarization; the QRS complex represents ventricular depolarization; and the T wave corresponds to ventricular repolarization. The R-R interval is the standard metric for measuring the time between heartbeats.

delay introduced by the atrioventricular (AV) node. After the electrical impulse exits the AV node, it is transmitted to the ventricles, triggering ventricular systole. This event is visible on the ECG as a series of three waves known as Q, R, and S, commonly referred to collectively as the QRS complex. Ventricular systole, and consequently the QRS complex, is brief due to the rapid conduction of the Purkinje system, and exhibits a larger amplitude compared to the P wave due to the greater number of cells in the ventricles relative to the atria. The final waveform, the T wave, is produced by repolarization of the ventricles during diastole. Electrodes placed on the chest, arms, and legs enable the detection of these waves, a technique known as electrocardiography (ECG). This electrode distribution permits the separate analysis of various combinations, referred to as leads, which record the heart's electrical activity along different spatial axes, enabling noninvasive monitoring of physiological and pathological patterns. Consequently, the ECG is considered the gold standard for diagnosing conditions such as arrhythmias, heart blocks, and myocardial infarction.

Due to the transient nature of certain conditions, such as arrhythmias, specialists often prescribe the use of portable ECG devices known as Holter monitors. These devices record cardiac activity over long periods. Despite their lightweight design and use of fewer leads, prolonged usage is not advised due to potential skin irritation caused by the electrodes,

which prevents screening applications. Moreover, even with ongoing miniaturization, these devices can be uncomfortable, primarily because of the cables that transmit signals from the electrodes to the device. Some manufacturers have begun producing single-lead Holter monitors, such as the Bittium MiniSnap, which eliminate these cables. As an alternative to Holter-type devices, there are also approaches such as the armband proposed in [17], capable of long-term ECG recordings.

In the context of wearable technology, chest straps and smartwatches are the most widely accepted devices [18, 19], both utilizing single-lead ECGs. Chest straps employ two electrodes placed on the chest, held by an elastic band, with signals transmitted wirelessly to another device, such as a smartphone or smartwatch. These signals are recorded throughout the duration of wear; however, prolonged use may lead to irritation, restricting their use primarily to exercise contexts. Smartwatches likewise use single-lead ECGs, utilizing two electrodes: one positioned on the back of the watch in contact with the skin, and the other on the crown, where the opposite hand's fingers are placed [20]. This is equivalent to lead I of a standard ECG. Within this lead, the QRS complex, indicative of ventricular systole, can be seen. The ventricles contract involving numerous cells almost instantaneously, resulting in a high-energy spike wave. This unique morphological and energetic profile significantly aids in the automatic identification of heartbeats.

Due to their comfort, smartwatches are suitable for longer periods of wear but cannot record ECG continuously. This limitation is due to the need to touch the smartwatch with both hands, since the electric loop must be closed, restricting recordings to short durations. Although enabling individuals to record ECG data for the evaluation of palpitations, syncope, and sudden death risk is promising, its effectiveness must still be thoroughly validated [21]. One way to overcome this limitation is through synergy with other continuous monitoring signals, such as PPG. For arrhythmias, PPG might offer sensitive markers that, although not very specific, could encourage users to conduct an ECG. Another drawback of smartwatches is the single-lead limitation, which may be sufficient for sports and fitness applications, but remains a limitation for healthcare purposes. Nonetheless,

studies have demonstrated the feasibility of using smartwatches to record both frontal and precordial leads [22], potentially expanding their clinical applications.

2.3 Peripheral venous pressure

Peripheral venous pressure (PVP) is a measurement of the blood pressure within the peripheral veins, typically measured in the limbs. This parameter is important in assessing the fluid status and venous return to the heart. PVP is often used in clinical settings as an indirect indicator of central venous pressure (CVP, see Figure 4), which reflects right atrial pressure and provides insight into cardiac function and intravascular volume status. Its uses include managing blood volume in scenarios of dehydration, hemorrhage, or heart failure; examining venous blockages and thrombosis; and providing guidance for intravenous treatment [23, 24].

As venous pressure is much lower compared to arterial pressure, precise placement of the piezoelectric sensor over the vein is necessary. Some methods incorporate the use of ultrasound alongside piezoelectric sensors to image the veins, a technique known as compression sonography [25]. In addition, PVP waveforms can be affected by the respiratory cycle. During inspiration, negative intrathoracic pressure can reduce venous pressure, while expiration can increase it. This respiratory variation can be particularly pronounced in PVP measurements. The compliance of the venous walls can also affect the waveform. Increased venous compliance can result in a more flattened waveform, while decreased compliance (e.g., due to venous congestion) may produce a sharper waveform.

PVP is typically measured using a catheter inserted into a peripheral vein, such as the cephalic or basilic vein in the arm. Recent studies have explored noninvasive methods to measure venous pressure using piezoelectric sensors [27, 28], with the aim of providing continuous monitoring without the risks associated with invasive catheterization. In this thesis, the PVP signal is acquired using a wrist-worn device known as NIVA, which stands for noninvasive Venous Waveform Analysis, engineered by VoluMetrix

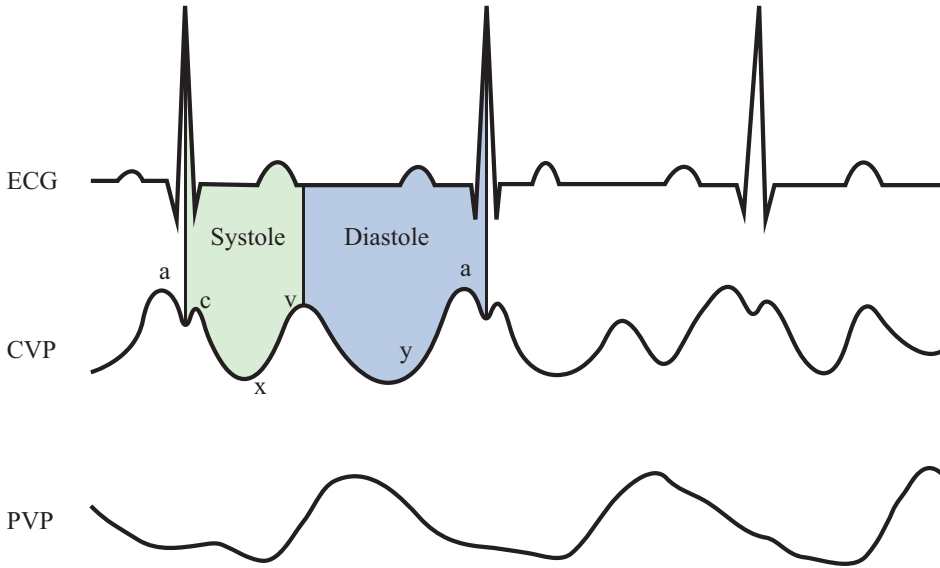


Figure 4: Comparison of CVP and PVP and relation with the ECG. CVP is composed of several waves: a (atrial contraction), c (tricuspid valve elevation into the right atrium), x (downward movement of the contracted right ventricle), v (back-pressure wave from blood filling the right atrium) and y (tricuspid valve opens in early ventricular diastole). These waves are not present in PVP, which has a pulsatile waveform similar to the PPG.

(Nashville, TN). The development of this device was motivated by some recent results of peripheral intravenous analysis that showed it is a promising indicator in blood volume assessment and hemorrhages [29, 30]. In Chapter 5 it is investigated how this signal can provide respiratory-guided parasympathetic markers.

2.4 Auxiliar signals

The signals discussed below are not the main focus of this thesis but have been used in various stages of the methodology. Although these signals can be integrated with data from wearable devices, they are typically not included in the sensor arrays of the most common wearables. These signals are utilized either for verification purposes or as a provisional solution until it becomes possible to depend solely on signals from wearable devices.



Figure 5: NIVA device. This device designed by Volumetrix (Nashville, TN) measure the changes in PVP in the wrist using a piezoelectric sensor. From [26].

Respiratory effort

The respiratory effort signal is obtained by a plethysmography band placed on the chest. When breathing in, the chest widens, stretching the band. The increased tension on the sensor is reflected as an increase in the signal. The opposite occurs on expiration. It should be noted that the movement of the chest does not necessarily indicate that air is entering or leaving the lungs. In diseases such as obstructive sleep apnea, the patient may make breathing movements at the torso level, but the airflow may be interrupted by the collapse of the upper airway.

Here, an application of this signal is evident, as it is crucial to determine whether an apnea is caused by obstruction (obstructive apnea) or by a lack of respiratory effort (central apnea), which are differentiated processes. Breathing is an involuntary action controlled by the autonomic nervous system. This signal also provides information about this system.

Nasal pressure

Nasal pressure is among the signals monitored during polysomnography, the test used to diagnose sleep apnea. It assesses the air pressure entering and exiting the nose during breathing through a sensor placed in a cannula. Under normal breathing, an oscillatory signal similar to the respiratory effort is recorded. However, this signal diminishes in amplitude or vanishes entirely during apneic events. This signal is crucial in identifying the cessation of airflow during apnea, while respiratory effort helps to distinguish between obstructive and central events, as previously mentioned.

3 Physiology

3.1 Autonomic nervous system

The autonomic nervous system (ANS) is a division of the nervous system responsible for controlling most visceral functions of the body, including the heart rate, blood pressure, sweating, and gastrointestinal motility. Its primary role is to assist in maintaining a stable internal environment, a process referred to as homeostasis [31]. The ANS provides sensory input to the central nervous system (CNS), which includes the brain and spinal cord, regarding the status of the internal organs. It is responsible for evoking sensations such as hunger, thirst, and nausea, which facilitate the maintenance of homeostasis through voluntary behaviors like drinking or eating. Although the ANS can evoke these perceptions and behaviors, its primary regulation occurs unconsciously. Internal stimuli, detected by various sensors within the viscera, trigger compensatory mechanisms that are primarily mediated by centers in the spinal cord, brainstem, and hypothalamus. The ANS can also function through visceral reflexes, which are neural connections that do not require processing by the brain, thus enabling rapid, subconscious reflex responses [32].

Sympathetic and parasympathetic divisions

The signals that return to the organs, known as efferent signals, are transmitted by the ANS through two subdivisions: sympathetic and parasympathetic. Both motor divisions communicate with the organs through a two-neuron pathway. The first neuron, known as the preganglionic neuron, has its cell body located within the CNS, while the second neuron, known as the postganglionic neuron, resides in an autonomic ganglion and is connected to smooth muscle, cardiac muscle, glands, or walls of the gastrointestinal tract, as shown in Figure 6. It is important to note that not all tissues receive innervation from both divisions; for example, most blood vessels, glands, and smooth muscle are primarily influenced by the sympathetic division. Although sympathetic and parasympathetic stimuli can induce both inhibitory and stimulatory effects, as demonstrated in Table 1, they generally exert antagonistic effects in the same organ.

The sympathetic nervous system, often referred to as the "fight or flight" system, prepares the body for stressful or emergency situations by increasing heart rate, dilating airways to improve oxygen intake, dilating pupils, and redirecting blood flow from the gastrointestinal tract to muscles.

Table 1: Autonomic effects on various organs [32]. Sympathetic and parasympathetic branches usually have opposite effects on the same tissue.

Organ	Sympathetic effect	Parasympathetic effect
Pupil	Dilated	Constricted
Heart	Increased rate	Slowed rate
Blood vessels	Most often constricted	Most often little or no effect
Bronchi	Dilated	Constricted
Gut	Decreased peristalsis and tone	Increased peristalsis and tone
Liver	Glucose released	Slight glycogen synthesis
Blood	Increased coagulation	None
Skeletal muscle	Increased glycogenolysis and strenght	None
Kidney	Decreased urine output	None
Glands	Vasoconstriction and slight secretion	Stimulation of copious secretion
Sweat glands	Copious sweating	Sweating on palms of hands

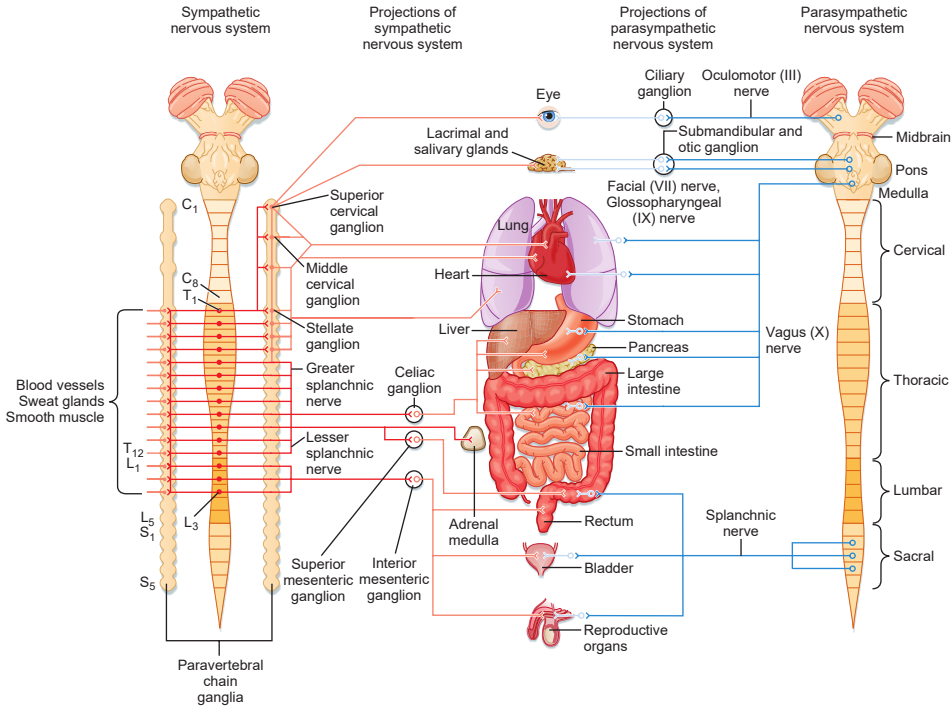


Figure 6: Sympathetic and Parasympathetic Pathways. From [31].

These responses, essential for rapid physical activity and heightened alertness, are mediated by neurotransmitters such as norepinephrine and epinephrine. In contrast, the parasympathetic nervous system, known as the "rest and digest" system, promotes relaxation and recovery by slowing the heart rate, constricting the airways, and stimulating digestive processes. The primary neurotransmitter involved is acetylcholine, which facilitates the restoration of the body to a calm state and supports energy conservation.

Although they often have antagonistic effects, the parasympathetic and sympathetic systems act in a coordinated manner to maintain internal balance. This coordination is exemplified in Table 1. During periods of rest, the predominance of parasympathetic tone ensures a low heart rate and relaxed breathing, while stimulating peristaltic movements in the intestine and other parts of the gastrointestinal tract, thus aiding digestion. In contrast, during a sprint to catch an arriving bus, sympathetic tone predominates, leading to bronchodilation and increased heart rate and contractility to meet

the increased demand for oxygen and blood in tissues. Simultaneously, the strength of skeletal muscle contractions is enhanced, and peristaltic movements are inhibited.

ANS interacts with the central nervous system, particularly the hypothalamus and brainstem, which integrate signals from various parts of the body and coordinate appropriate responses. These interactions ensure that physiological homeostasis is maintained, allowing the body to pose significant health challenges, affecting cardiovascular function, thermoregulation, and other critical systems. Understanding the autonomic nervous system is essential for developing treatments for a variety of conditions, ranging from heart failure to gastrointestinal disorders.

3.2 Autonomic control of the cardiorespiratory system

The cardiorespiratory system, comprising the heart, blood vessels, and lungs, is intricately regulated by the ANS, which ensures that the body meets its metabolic demands efficiently (see Figure 7). ANS is crucial in modulating heart rate, blood pressure, and respiratory rate to maintain homeostasis and respond to varying levels of physical activity or stress. The sympathetic nervous system increases heart rate and cardiac contractility, which boosts cardiac output and enhances blood flow to muscles and vital organs. This is mediated by the release of catecholamines, such as norepinephrine and epinephrine, which bind to adrenergic receptors in the heart and vasculature. In addition, the sympathetic nervous system causes bronchodilation, widening the airways to improve oxygen uptake, and increasing respiratory rate to facilitate greater oxygen delivery to the bloodstream. In contrast, the parasympathetic nervous system reduces heart rate and decreases cardiac output through the action of acetylcholine on muscarinic receptors in the heart. The parasympathetic nervous system primarily influences heart rate through the vagus nerve, which slows heart rate and helps maintain a resting state. Its effect on the respiratory system is less direct but generally promotes a more relaxed breathing pattern, consistent with states of rest. The balance between sympathetic and parasympathetic activity is vital for

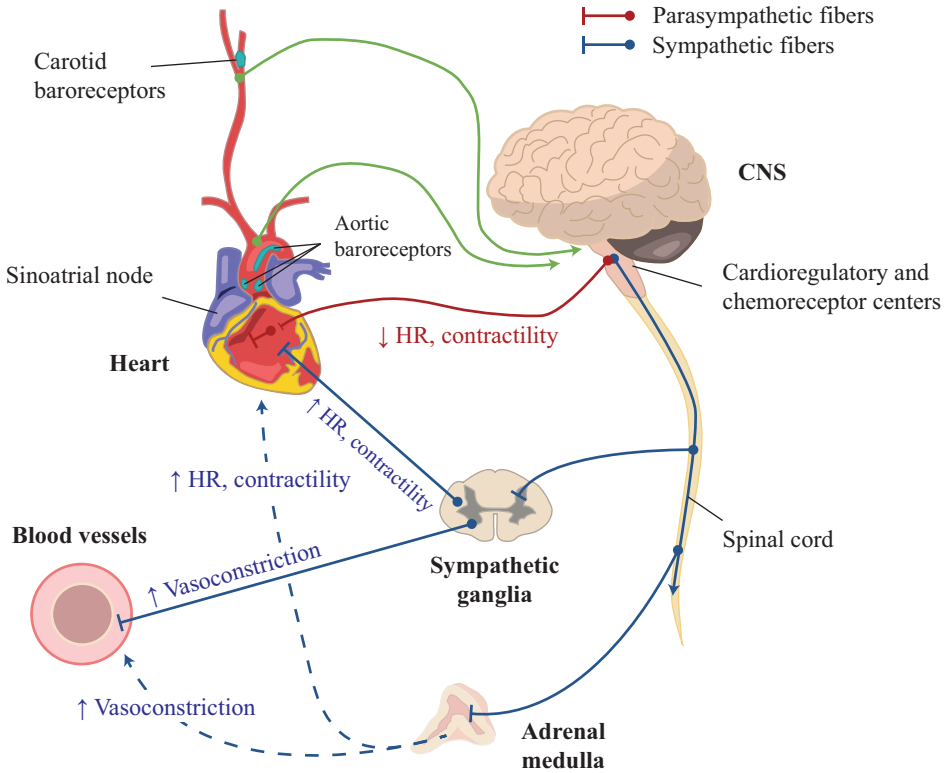


Figure 7: Illustration of the short-term ANS regulatory mechanisms of the cardiovascular system. Vectorized and adapted from [39].

maintaining cardiovascular and respiratory homeostasis. During exercise or stress, the sympathetic nervous system dominates, ensuring that adequate oxygen and nutrients are delivered to muscles and organs by increasing heart rate and respiratory rate. In contrast, during rest, the parasympathetic nervous system predominates, reducing these rates to conserve energy and maintain baseline bodily functions. Dysregulation of the ANS can lead to various cardiovascular and respiratory disorders. Conditions such as hypertension [33], heart failure [34, 35], and arrhythmias [36] can be the result of imbalances in autonomic control. Similarly, autonomic dysfunction is related to conditions such as sleep apnea [37] and mood disorders [38].

Control of the heart

Certain cells within heart tissue possess the ability to become self-excited, a property primarily concentrated in the sinoatrial (SA) node, also known as the sinus node, located in the wall of the right atrium (see Figure 8). The SA node comprises a group of cells with a high degree of automatism that, under normal physiological conditions, regulate the cardiac rhythm, functioning as the heart's natural pacemaker. The self-excitation of SA node cells generates electrical impulses in a rhythmic manner, which are transmitted through rapid conduction pathways across the atria. These electrical impulses, called action potentials, induce atrial muscle contraction, a process known as atrial systole, during which blood is pumped from the atria into the ventricles. At this juncture, the action potential reaches the AV node, where it experiences a brief delay, allowing for complete atrial emptying. Following this delay, the action potential rapidly propagates through the ventricles via the Purkinje system, initiating ventricular systole, which propels blood out of the heart and into the tissues. The final phase of the cardiac cycle is diastole, characterized by the relaxation of the heart muscle and repolarization of the cells, preparing them to respond to the next action potential.

Although the heart is capable of autonomously pumping blood, even continuing to beat outside the body due to its self-excitatory capability, the body does not require a uniform blood supply throughout the day, necessitating external regulation. For example, during emotional arousal or physical exercise, the heart rate increases, while it decreases during sleep. As illustrated in Figure 8, the heart is innervated by both the sympathetic and parasympathetic divisions of the autonomic nervous system. The parasympathetic division, primarily through the vagus nerve, innervates the SA and AV nodes, reducing self-excitation of the SA node cells and thereby decreasing heart rate. It also reduces the excitability of the AV node, prolonging the delay at this junction and extending the duration of the heartbeat. In contrast, sympathetic division exerts opposing effects: it increases the heart rate by stimulating the SA node and enhances the excitability of all cardiac cells, facilitating the transmission of electrical

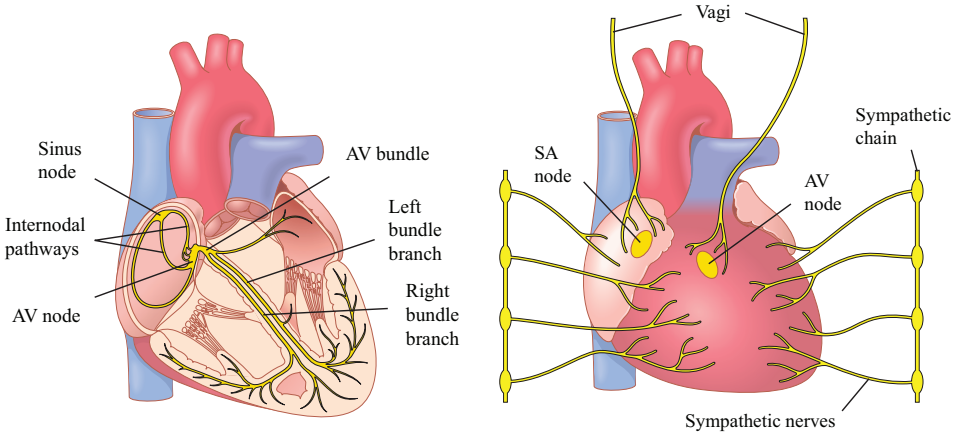


Figure 8: Cardiac conduction and innervation. From [32]. *Left:* Electrical conduction pathways. *Right:* Autonomic innervation.

impulses. In addition, it increases the force of heart muscle contraction.

Both types of autonomic stimulation can act rapidly and produce distinct effects: the parasympathetic nervous system is capable of completely halting the rhythmic excitation of the SA node, while the sympathetic nervous system can triple the normal heart rate and double the force of contraction [32]. Activation of the sympathetic nervous system results in the release of norepinephrine, which binds to beta-adrenergic receptors in the SA node. This interaction accelerates the depolarization rate of the nodal cells, resulting in an increased heart rate. In contrast, the parasympathetic nervous system decreases heart rate by releasing acetylcholine from the vagus nerve. Acetylcholine binds to muscarinic receptors in the SA node, causing hyperpolarization of the nodal cells. This hyperpolarization reduces the depolarization rate, thereby resulting in a slower heart rate.

3.3 Heart rate variability

Earlier sections discussed methods for acquiring biosignals related to the cardiorespiratory system, as well as how these signals are regulated by the ANS. Consequently, it can be inferred that analyzing these signals enables the extraction of information concerning the ANS. This presents a significant

advantage, as the direct acquisition of signals from the ANS is complex and, at present, cannot be achieved in a noninvasive way. This discussion focuses on heart rate variability (HRV), a powerful technique that leverages the regulatory influence of the ANS on the SA node.

Oscillatory patterns in heart rate and blood pressure, along with their correlation with the respiratory cycle, have been recognized since the 18th century [40]. In the 19th century, cyclical changes in arterial pressure, known as 10-second waves, were discovered by Mayer [41]. However, the broad clinical significance of heart rate oscillations around a mean heart rate was not fully realized until the 1960s, with the advent of computer technology. Numerous studies have demonstrated that autonomic monitoring through HRV has the potential to predict major bodily disorders, some of which are life-threatening, that can be prevented using minimally invasive devices and techniques. HRV reflects the integrated response of the cardiovascular system to various influences [42], and has also contributed to elucidating previously poorly understood mechanisms of autonomic control.

Under normal conditions, a healthy heart does not maintain a constant interval between beats; instead, the heart rate continuously fluctuates in response to various internal and external stimuli. HRV reflects these fluctuations in the intervals between consecutive heartbeats and serves as an indicator of the heart's capacity for short-term adaptation. Furthermore, HRV has been shown to be a valuable tool for the prognosis, diagnosis, and monitoring of diseases [37, 42–47]. HRV has proven useful in a wide range of conditions, including myocardial infarction [42], diabetic neuropathy [43], essential hypertension [44], ischemic heart disease, congestive heart failure, Chagas disease [45], monitoring of cardiac transplant recipients [46], and even depression [47].

Typically, HRV is derived from the ECG. HRV can be assessed by identifying the P waves in an ECG recording and calculating the P-P intervals, *i.e.*, the time intervals between consecutive heartbeats. Nevertheless, due to the relative simplicity associated with the detection of QRS complexes, the beat-to-beat interval is often determined from the R-R interval in practical applications. HRV analysis can be performed using temporal, frequential,

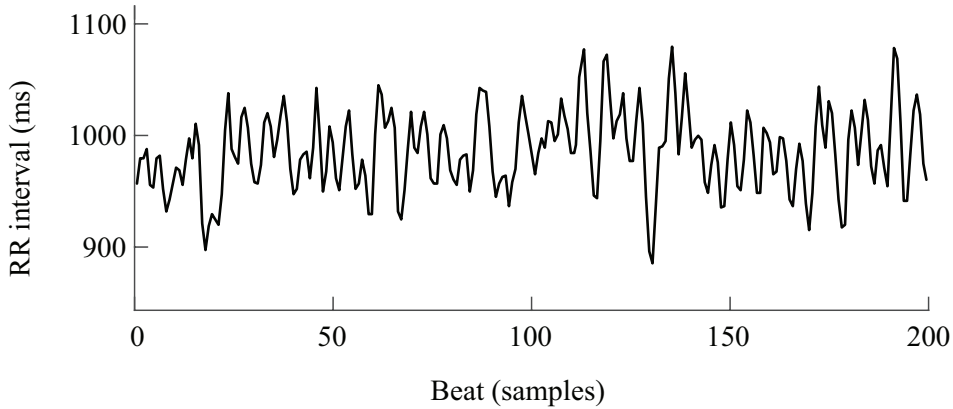


Figure 9: Example of a tachogram. The R-R interval physiologically fluctuates around a mean heart rate.

and nonlinear methods, each offering unique information on how the autonomic nervous system modulates cardiac function. Although time and frequency domain metrics have traditionally been the focus of research, newer approaches, such as Poincaré plots, have become integral to HRV analysis, providing prognostic information in conditions such as myocardial infarction, chronic heart failure, and sudden infant death syndrome [48].

To obtain HRV indices, it is necessary to calculate the R-R intervals as the difference between two consecutive R waves, *i.e.*, the inverse of the instantaneous heart rate. The representation of this signal is called the tachogram (see Figure 9), where oscillations around a mean value are clearly depicted.

Variables such as the mean heart rate can be directly derived from the tachogram. The variance, as a measure, reflects the total fluctuations in heart rate around the mean, thereby representing the total power of HRV, regardless of its origin. Despite this limitation, simple time domain parameters remain valuable for evaluating HRV. Sympathetic activation typically results in tachycardia, which is usually accompanied by a significant reduction in total power, while vagal activation produces the opposite effect [37]. Tachycardia is observed as an increase in mean heart rate, *i.e.*, a shortened R-R interval, while the corresponding reduction in total power is evidenced by a decrease in variance. In a study by [46], cardiac transplant

recipients, whose hearts are non-innervated, exhibited a 96% reduction in HRV, underscoring the role of the ANS in modulating heart rate fluctuations. This finding implies that HRV is primarily an indicator of neural modulation by the ANS.

More informative variables can be observed in the frequency domain. Several methods can be employed to estimate the power spectral density, either directly from the event series, which are intrinsically unevenly sampled, or through interpolation to obtain an evenly sampled signal. According to the Nyquist-Shannon sampling theorem, this spectrum cannot contain information above half the sampling rate, which in this case is the heart rate as the information is sampled beat-to-beat. Three components can be distinguished: a very low frequency component (VLF, < 0.04 Hz); a low frequency component (LF, $0.04 - 0.15$ Hz), which is usually related to the activation of the sympathetic and parasympathetic divisions of the ANS; and a high frequency component (HF, $0.15 - 0.4$ Hz), related to the parasympathetic division. The latter corresponds mainly to the frequencies of respiratory sinus arrhythmia (RSA, *i.e.*, shortened R-R intervals during inspiration and prolonged during expiration), although, depending on the respiratory rate, these bounds may not be precise, and special methods should be applied (see Chapter 4).

The significance and association of these components with physiological processes have been extensively studied since the 1980s, following the foundational work of Akselrod *et al.* in 1981 [49]. In their research, they demonstrated, through the monitoring of selective autonomic blockade effects in dogs, that the HF component is related to parasympathetic modulation, whereas the LF component is associated with both sympathetic and parasympathetic activity, as well as the renin-angiotensin system. This novel approach enabled the differentiation of modulations attributable to each motor division of the ANS noninvasively, a distinction previously achievable only through invasive methods [45]. Consistent findings were also reported in humans [50], indicating that the LF component of HRV increases from supine to standing positions, suggesting a connection between postural changes and autonomic control mediated by blood pressure and

baroreflex mechanisms within this frequency range. In addition, controlled respiration was shown to enhance the HF component, demonstrating its relationship with respiration. This increase in the HF component, accompanied by a significant reduction in the LF component, was also observed by Pagani et al. [51].

Later in the same decade, Sands et al. observed that in cardiac transplant recipients, the prominent peak at the respiratory rate, *i.e.*, the HF component, was either absent or substantially diminished, suggesting that vagal innervation is essential for RSA [46]. Despite extensive research, the physiological significance of the VLF band remains unclear, although it has been suggested that it is associated with cyclic fluctuations in peripheral vasomotor tone related to thermoregulation [52]. Moreover, achieving precise spectral estimation in this frequency band demands extended recordings, during which the signal might not stay stationary, adding complexity to the analysis.

These studies have led to the following conclusions regarding HRV: 1) the respiratory oscillations, identified as the HF component, serve as a marker of parasympathetic (vagal) modulation; 2) the rhythm corresponding to vasomotor waves, identified as the LF component, serves as a marker of both sympathetic and parasympathetic modulations; and 3) these rhythms exhibit a reciprocal relationship analogous to that characterizing sympathovagal balance [45].

In addition to time and frequency-domain metrics, non-linear and symbolic metrics have been developed in recent years to uncover hidden information in HRV that traditional methods may not reveal. HRV metrics are discussed in detail in Chapter 4.

Variability measurements using wearables

Despite advances in technology, HRV measurement from the ECG still requires the use of contact electrodes, such as adhesive strips or textile-based sensors. This requirement limits their practicality and acceptance for daily monitoring, particularly among nonpatient populations. Alternatively, PPG allows for the monitoring of pulse rate variability (PRV) by detecting pulse

signals. Although PRV differs from HRV, it can serve as a substitute in many practical applications [53, 54].

Several commercial wristband wearables claim to measure PRV using PPG on the wrist, although few studies have validated these measurements against ECG-derived HRV [55], revealing significant limitations in the robust estimation of variability metrics. Therefore, there is a clear need in the market for noninvasive systems that provide reliable HRV/PRV estimates at low cost and are widely accepted by the general population. In addition, numerous wrist devices offer heart rate information; however, while several studies have validated PRV obtained from PPG signals recorded on the finger and forehead [56], very few have validated PRV from wrist-based PPG during daily life. Most studies that extract heart rhythm information from wrist-based PPG signals focus only on mean heart rate. However, accurate HRV analysis requires the detection of each individual pulse, a task that is often compromised in ambulatory conditions. The proliferation of wearable devices necessitates further investigation of the degradation of HRV metrics due to incomplete recordings, where pulse detections may be lost due to noise and movement.

Bibliography

- [1] Yoonhyuk Jung, Seongcheol Kim, and Boreum Choi. Consumer valuation of the wearables: The case of smartwatches. *Computers in Human Behavior*, 63:899–905, 10 2016.
- [2] Wen Yen Lin, Wen Cheng Chou, Tsai Hsuan Tsai, Chung Chih Lin, and Ming Yih Lee. Development of a wearable instrumented vest for posture monitoring and system usability verification based on the technology acceptance model. *Sensors (Switzerland)*, 16(12), 12 2016.
- [3] Phaik Khee Beh, Yuvaraj Ganesan, Mohammad Iranmanesh, and Behzad Foroughi. Using smartwatches for fitness and health monitoring: the UTAUT2 combined with threat appraisal as moderators. *Behaviour and Information Technology*, 40(3):282–299, 2021.
- [4] Dariusz Mrozek, Anna Koczur, and Bożena Małysiak-Mrozek. Fall detection in older adults with mobile IoT devices and machine learning in the cloud and on the edge. *Information Sciences*, 537:132–147, 10 2020.
- [5] Man Lai Cheung, Wilson K.S. Leung, and Haksin Chan. Driving healthcare wearable technology adoption for Generation Z consumers in Hong Kong. *Young Consumers*, 22(1):10–27, 2020.
- [6] Tsai Hsuan Tsai, Wen Yen Lin, Yung Sheng Chang, Po Cheng Chang, and Ming Yih Lee. Technology anxiety and resistance to change behavioral study of a wearable cardiac warming system using an extended TAM for older adults. *PLoS ONE*, 15(1), 1 2020.

- [7] Yiwen Gao, He Li, and Yan Luo. An empirical study of wearable technology acceptance in healthcare. *Industrial Management and Data Systems*, 115(9):1704–1723, 10 2015.
- [8] Krishnan Gopinath, Gowthaman Selvam, and Gopalakrishnan Narayanamurthy. Determinants of the Adoption of Wearable Devices for Health and Fitness: A Meta-analytical Study. *Communications of the Association for Information Systems*, 50(1):445–480, 2022.
- [9] Muhammad Shabaan, Kaleem Arshad, Muhammad Yaqub, Fung Jinchao, M. Sultan Zia, Girridhar Reddy Boja, Muazzam Iftikhar, Usman Ghani, Loknath Sai Ambati, and Rizwan Munir. Survey: Smartphone-based assessment of cardiovascular diseases using ECG and PPG analysis, 7 2020.
- [10] Leif Sörnmo and Pablo Laguna. *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Elsevier, 2005.
- [11] Lloyd Price. Physics and Light : Apple’s competitive advantage in Healthcare, 4 2020.
- [12] Jack Sexty. Polar launch the OH1+ armband heart rate monitor, with Bluetooth and ANT+ connectivity.
- [13] Rozalynn S. Frazier. Why Some People Shouldn’t Use Fitness Trackers, According to Experts.
- [14] Ketut Subiyanto. Crop fit sportswoman checking fitness tracker in nature.
- [15] Kento Watanabe, Shintaro Izumi, Yuji Yano, Hiroshi Kawaguchi, and Masahiko Yoshimoto. A 5-ms Error, 22- μ A Photoplethysmography Sensor using Current Integration Circuit and Correlated Double Sampling. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5566–5569. IEEE, 7 2018.
- [16] Bojana Koteska, Ana Madevska Bodanova, Hristina Mitrova, Marija Sidorenko, and Fedor Lehocki. A Deep Learning Approach to Estimate

- SpO₂ from PPG Signals. In *Proceedings of the 9th International Conference on Bioinformatics Research and Applications*, pages 142–148, New York, NY, USA, 9 2022. ACM.
- [17] Jesus Lazaro, Natasa Reljin, Md Billal Hossain, Yeonsik Noh, Pablo Laguna, and Ki H. Chon. Wearable Armband Device for Daily Life Electrocardiogram Monitoring. *IEEE Transactions on Biomedical Engineering*, 67(12):3464–3473, 12 2020.
- [18] C. Peng, N. Xi, H. Zhao, and J. Hamari. Acceptance of Wearable Technology: A Meta-Analysis. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, volume 2022-Janua, pages 5101–5110, 2022.
- [19] M. Simonnet, B. Gourvennec, and R. Billot. Connected Heart Rate Sensors to Monitor Sleep Quality: Electrodes, Chest Belt and Smartwatch Users Acceptability. In *Proceedings - 2016 IEEE 1st International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2016*, pages 344–345, 2016.
- [20] Nino Isakadze and Seth S. Martin. How useful is the smartwatch ECG?, 10 2020.
- [21] Marc Strik, Sylvain Ploux, F. Daniel Ramirez, Saer Abu-Alrub, Pierre Jaïs, Michel Haïssaguerre, and Pierre Bordachar. Smartwatch-based detection of cardiac arrhythmias: Beyond the differentiation between sinus rhythm and atrial fibrillation, 9 2021.
- [22] Joske van der Zande, Marc Strik, Rémi Dubois, Sylvain Ploux, Saer Abu Alrub, Théo Caillol, Mathieu Nasarre, Dirk W. Donker, Eline Oppersma, and Pierre Bordachar. Using a Smartwatch to Record Pre-cordial Electrocardiograms: A Validation Study. *Sensors*, 23(5), 3 2023.
- [23] Kazuya Nagao, Shiori Maruichi-Kawakami, Kenji Aida, Kenichi Matsuto, Kazumasa Imamoto, Akinori Tamura, Tadashi Takazaki, Taro Nakatsu, Masaru Tanaka, Shogo Nakayama, Takeshi Morimoto,

- Takeshi Kimura, and Tsukasa Inada. Association of peripheral venous pressure with adverse post-discharge outcomes in patients with acute heart failure: a prospective cohort study. *European Heart Journal. Acute Cardiovascular Care*, 11(5):407–417, 6 2022.
- [24] Olivier Desebbe, Sylvain Vallier, Laurent Gergelé, Brenton Alexander, Alexandre Marx, Elias Ben Jaoude, Hiromi Kato, Leila Toubal, Antoine Berna, Jacques Duranteau, Jean-Louis Vincent, and Alexandre Joosten. Diagnostic accuracy of the peripheral venous pressure variation induced by an alveolar recruitment maneuver to predict fluid responsiveness during high-risk abdominal surgery. *BMC Anesthesiology*, 23(1):249, 7 2023.
- [25] Christopher M. Hearon, Kirsten Peters, Katrin A. Dias, James P. Macnamara, John E. T. Marshall, Joseph Campaign, David Martin, Karina Marshal-Goebel, and Benjamin D. Levine. Assessment of venous pressure by compression sonography of the internal jugular vein during 3 days of bed rest. *Experimental Physiology*, 108(12):1560–1568, 12 2023.
- [26] DAIC Heart Failure. VoluMetrix’s NIVA|HF for Heart Failure Patients Designated as Breakthrough Device by U.S. FDA, 6 2021.
- [27] Bret D. Alvis, Monica Polcz, Jessica H. Huston, Timothy S. Hopper, Phil Leisy, Kelly Mishra, Susan S. Eagle, Colleen M. Brophy, Joann Lindendorf, and Kyle M. Hocking. Observational Study of Noninvasive Venous Waveform Analysis to Assess Intracardiac Filling Pressures During Right Heart Catheterization. *Journal of Cardiac Failure*, 26(2):136–141, 2 2020.
- [28] Sarah Flint, Aaron Wengrofsky, Peter P Vlismas, Elliot Wiesenfeld, Yogita M Rochlani, Patricia chavez, Sasha Vukelic, Sandhya Murthy, Shivank Madan, Omar Saeed, Snehal R Patel, Jooyoung Shin, Ulrich P Jorde, and Daniel B Sims. Abstract 14769: Peripheral Venous Pressure Measured With a Novel Miniature Pressure Transducer Correlates With Central Venous Pressure in Heart Failure, Heart Transplant, and

- Left Ventricular Assist Device Patients. *Circulation*, 148(Suppl_1), 11 2023.
- [29] Kyle M. Hocking, Ban Sileshi, Franz J. Baudenbacher, Richard B. Boyer, Kelly L. Kohorst, Colleen M. Brophy, and Susan S. Eagle. Peripheral Venous Waveform Analysis for Detecting Hemorrhage and Iatrogenic Volume Overload in a Porcine Model. *Shock*, 46(4):447–452, 10 2016.
- [30] K.M. Hocking, B.D. Alvis, F. Baudenbacher, R. Boyer, C.M. Brophy, I. Beer, and S. Eagle. Peripheral i.v. analysis (PIVA) of venous waveforms for volume assessment in patients undergoing haemodialysis. *British Journal of Anaesthesia*, 119(6):1135–1140, 12 2017.
- [31] Bruce M Koeppen and Bruce A Stanton. Berne and Levy Physiology: Berne and Levy Physiology E-Book. *Elsevier Health Sciences*, 2017.
- [32] SeyedAli Khonsary. Guyton and Hall: Textbook of Medical Physiology. *Surgical Neurology International*, 8(1):275, 2017.
- [33] Solomon Tesfaye, Nish Chaturvedi, Simon E.M. Eaton, John D. Ward, Christos Manes, Constantin Ionescu-Tirgoviste, Daniel R. Witte, and John H. Fuller. Vascular Risk Factors and Diabetic Neuropathy. *New England Journal of Medicine*, 352(4):341–350, 1 2005.
- [34] Maria Teresa La Rovere, Gian Domenico Pinna, Roberto Maestri, Andrea Mortara, Soccorso Capomolla, Oreste Febo, Roberto Ferrari, Mariella Franchini, Marco Gnemmi, Cristina Opasich, Pier Giorgio Riccardi, Egidio Traversi, and Franco Cobelli. Short-Term Heart Rate Variability Strongly Predicts Sudden Cardiac Death in Chronic Heart Failure Patients. *Circulation*, 107(4):565–570, 2 2003.
- [35] Patrice G. Guyenet. The sympathetic control of blood pressure. *Nature Reviews Neuroscience*, 7(5):335–346, 5 2006.
- [36] Yu-ki Iwasaki, Kunihiro Nishida, Takeshi Kato, and Stanley Nattel. Atrial Fibrillation Pathophysiology. *Circulation*, 124(20):2264–2274, 11 2011.

- [37] Task Force of The European Society of Cardiology and the North American Society for Pacing and Electrophysiology. Heart rate variability Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3):354–381, 1996.
- [38] Carmen Schiweck, Deborah Piette, Daniel Berckmans, Stephan Claes, and Elske Vrieze. Heart rate and high frequency heart rate variability during stress as biomarker for clinical depression. A systematic review. *Psychological Medicine*, 49(2):200–211, 1 2019.
- [39] Marta Carrara, Manuela Ferrario, Bernardo Bollen Pinto, and Antoine Herpain. The autonomic nervous system in septic shock and its role as a future therapeutic target: a narrative review. *Annals of Intensive Care*, 11(1), 12 2021.
- [40] Stephen Hales. *Statistical essays*, volume 2. Gale, 1733.
- [41] Sigmund Mayer. Studien zur Physiologie des Herzens und der Blutgefäße V. über spontane Blutdruck schwankungen. *Akad Wiss Wien Math Nat Kl*, 74:281–307, 1876.
- [42] Gian Carlo Casolo, Piero Stroder, Claudia Signorini, Francesca Calzolari, Mauro Zucchini, Enrico Balli, Antonio Sulla, and Stefano Lazzarini. Heart Rate Variability During the Acute Phase of Myocardial Infarction. *Circulation*, 85(6):2073–2079, 1992.
- [43] Massimo Pagani, Gabriella Malfatto, Simona Pierini, Rodolfo Casati, Anna Maria Masu, Massimo Poli, Stefano Guzzetti, Federico Lombardi, Sergio Cerutti, and Alberto Malliani. Spectral analysis of heart rate variability in the assessment of autonomic diabetic neuropathy. *Journal of the Autonomic Nervous System*, 23:143–153, 1988.
- [44] Stefano Guzzetti, Emanuela Piccaluga, Rodolfo Casati, Sergio Cerutti, Federico Lombardi, Massimo Pagani, and Alberto Malliani. Sympathetic predominance an essential hypertension: a study employing spectral analysis of heart rate variability. *Journal of Hypertension*, 6(9):711–717, 1988.

-
- [45] Alberto Malliani, Massimo Pagani, Federico Lombardi, and Sergio Cerutti. Cardiovascular neural regulation explored in the frequency domain. *Circulation*, 84(2):482–492, 1991.
- [46] Kenneth E F Sands, Marvin L Appel, Leonard S Lilly, Frederick J Schoen, Gilbert H Mudge, and Richard J Cohen. Power Spectrum Analysis of Heart Rate Variability in Human Cardiac Transplant Recipients. *Circulation*, 79(1):76–82, 1989.
- [47] Thomas Rechlin, Maria Weis, Alfred Spitzer, and Wolfgang P Kaschka. Are affective disorders associated with alterations of heart rate variability? *Journal of Affective Disorders*, 32:271–275, 1994.
- [48] Berik Koichubekov, Viktor Riklefs, Marina Sorokina, Ilya Korshukov, Lyudmila Turgunova, Yelena Laryushina, Ryszhan Bakirova, Gulmira Muldaeva, Ernur Bekov, and Makhabbat Kultenova. Informative nature and nonlinearity of lagged poincaré plots indices in analysis of heart rate variability. *Entropy*, 19(10), 10 2017.
- [49] S. Akselrod, D. Gordon, F.A. Ubel, D.C. Shannon, A.C. Barger, and R.J. Cohen. Power spectrum analysis of heart rate fluctuation: A quantitative probe of beat-to-beat cardiovascular control. *Science*, 213(4504):220–222, 1981.
- [50] B. Pomeranz, J.B. Macaulay, and M.A. Caudill. Assessment of autonomic functions in humans by heart rate spectral analysis. *American Journal of Physiology - Heart and Circulatory Physiology*, 17(1), 1985.
- [51] M. Pagani, F. Lombardi, S. Guzzetti, O. Rimoldi, R. Furlan, P. Pizzinelli, G. Sandrone, G. Malfatto, S. Dell’Orto, and E. Piccaluga. Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circulation Research*, 59(2):178–193, 1986.
- [52] B.W. Hyndman. The role of rhythms in homeostasis. *Kybernetik*, 15(4):227–236, 1974.

-
- [53] E. Gil, M. Orini, R. Bailón, J. M. Vergara, L. Mainardi, and P. Laguna. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. *Physiological Measurement*, 31(9):1271–1290, 2010.
- [54] Axel Schäfer and Jan Vagedes. How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram. *International Journal of Cardiology*, 166(1):15–29, 6 2013.
- [55] David Hernando, Surya Roca, Jorge Sancho, Álvaro Alesanco, and Raquel Bailón. Validation of the apple watch for heart rate variability measurements during relax and mental stress in healthy subjects. *Sensors*, 18(8), 8 2018.
- [56] Elena Peralta, Jesus Lazaro, Raquel Bailon, Vaidotas Marozas, and Eduardo Gil. Optimal fiducial points for pulse rate variability analysis from forehead and finger photoplethysmographic signals. *Physiological Measurement*, 40(2), 2 2019.

1 Chronic mental health conditions

This thesis addresses techniques for monitoring the autonomic nervous system in obstructive sleep apnea and chronic mental health conditions. These diseases are very different from each other, but the common pattern is their long duration and slow progression. This temporal characteristic makes wearable devices very useful as screening and monitoring tools.

Chronic mental health conditions, such as depression, anxiety disorders, schizophrenia, and bipolar disorder, represent long-term psychological disturbances that persist for extended periods, often impacting individuals' daily functioning, quality of life, and social relationships. These conditions are characterized by their enduring nature and the recurrent or continuous symptoms that can fluctuate in intensity over time. In contrast to acute mental health episodes, which typically resolve within a brief time span, chronic mental health conditions necessitate continuous management and treatment to lessen their impact and help the individual maintain a fulfilling life. The complexity of chronic mental health conditions arises from their multifactorial etiology, involving a combination of genetic, environmental, psychological, and neurobiological factors. The management of these conditions often necessitates a multidisciplinary approach, including pharmacological interventions, psychotherapy, lifestyle modifications, and social

support systems. This thesis investigates the monitoring of patients with major depressive disorder and generalized anxiety disorder by measuring stress reactivity with wearable devices.

Major Depressive Disorder (MDD) is a prevalent and serious mental health condition characterized by persistent feelings of sadness, hopelessness, and a loss of interest or pleasure in most activities. These symptoms significantly impair daily functioning and can affect various aspects of life, including work, relationships, and physical health. MDD is more than just experiencing occasional low moods; it involves a deep and sustained depressive state that can last for weeks, months, or even longer if untreated. The diagnosis of MDD is based on specific criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). To meet these criteria, an individual must experience at least five of the following symptoms for a minimum of two weeks: depressed mood, diminished interest or pleasure in activities, significant weight change or appetite disturbance, sleep disturbances (insomnia or hypersomnia), psychomotor agitation or retardation, fatigue or loss of energy, feelings of worthlessness or excessive guilt, difficulty concentrating or making decisions, and recurrent thoughts of death or suicide [1]. The etiology of MDD is complex and multifactorial, involving genetic predisposition, neurobiological factors such as neurotransmitter imbalances (*e.g.*, serotonin, norepinephrine), hormonal influences, and psychosocial stressors. Despite its prevalence, MDD remains a leading cause of disability worldwide, with significant personal and societal burdens. Early diagnosis and effective treatment are critical in improving outcomes and reducing the risk of chronicity and relapse.

Generalized Anxiety Disorder (GAD) is a chronic mental health condition characterized by excessive, uncontrollable worry about a variety of everyday issues, such as health, work, social interactions, and routine life circumstances. This worry is disproportionate to the actual likelihood or impact of the feared events and is often accompanied by physical symptoms like restlessness, fatigue, difficulty concentrating, irritability, muscle tension, and sleep disturbances. These symptoms manifest on the majority of days over a period of at least six months, substantially impairing the individual's

capacity to function in daily activities. The diagnosis of GAD, according to DSM-5, requires the presence of excessive anxiety and worry that the individual finds difficult to control. Additionally, the anxiety must be associated with at least three of the following symptoms: restlessness or agitation, being easily fatigued, difficulty concentrating, irritability, muscle tension, and sleep disturbance [1]. The exact cause of GAD is not fully understood, but it is believed to involve a combination of genetic, neurobiological, and environmental factors. Additionally, individuals with a family history of anxiety disorders may be at a higher risk of developing GAD, suggesting a genetic predisposition. GAD is a common disorder that can be debilitating if left untreated. However, with appropriate intervention, many individuals can achieve significant relief of symptoms and improve their overall quality of life.

1.1 Relationship with stress reactivity

The human body functions as a dynamic system, constantly adjusting to its environment to maintain homeostasis. Interactions with the environment create various situations typically classified as either stressful or relaxing. Beyond the subjective feelings these situations produce, the ANS regulates energy expenditure during relaxation by decreasing heart rate, expanding blood vessels, and slowing breathing. The opposite processes occur in stressful or threatening situations. Although this physiological response is innate and shared by all mammals, providing an evolutionary advantage [2, 3], it can be affected by factors such as chronic stress, GAD, and MDD [3–7].

Numerous studies have documented diminished stress reactivity in individuals with depression and anxiety [3, 4, 6, 8–14]. This hyporeactivity is interpreted as a reduction in parasympathetic withdrawal and parasympathetic overactivation during stress, although the literature on the latter is inconsistent [4]. Some studies, however, indicate increased reactivity [7, 15–17], while others report both trends depending on the type of stressor [18, 19]. The causality of altered stress reactivity remains unclear: it may be a risk factor for developing GAD and MDD, or these disorders could

be the cause. Alternatively, a shared pathway might exist that includes GAD, MDD, and compromised stress reactivity [18, 20].

1.2 Wearable monitoring

Understanding the pathophysiology of GAD and MDD is rudimentary compared to other diseases, mainly due to the challenges in studying brain changes [21]. Evaluation of these disorders often relies on subjective assessments, including symptoms that frequently overlap between conditions [21]. Thus, it is crucial to investigate objective metrics to improve understanding and personalize treatment for depression, anxiety, and related disorders. Assessment of ANS through HRV provides objective data that complements other techniques, such as neuroimaging. High HRV levels at rest and greater reactivity are associated with attention and emotion processes that facilitate adaptive stress responses, better regulation, and executive function [3]. In addition, HRV can improve the understanding of interactions between various pathologies with established connections, such as the relationship between depression and sleep apnea or cardiovascular disease, which are also examined by HRV [3, 19, 22].

1.3 Main challenges

Although stress reactivity has gained significant traction in academic research, it remains a relatively unexplored field in clinical practice. This suggests that many challenges still need to be identified. For healthcare professionals to incorporate measurements from wearable devices, these devices must provide valuable and interpretable information that is applicable in clinical practice. Therefore, it is crucial to clarify which metrics are most appropriate and, importantly, when these measurements should be taken.

Wearable devices enable continuous monitoring, making periodic measurement feasible, provided the device has the capability to measure physiological variables uninterrupted. Throughout the day, individuals encounter various situations that activate the ANS to adapt, making this approach

useful for gathering as much information as possible. Recording all ANS states throughout the day would allow for a comprehensive analysis of stress reactivity, though it is important to note that not all observed states necessarily reflect a specific stress reactivity. Reactivity is best evaluated in response to specific events or in states of significant variability, allowing clearer response patterns to emerge.

This same variety of situations poses a difficulty in assessing whether changes in the ANS are appropriate responses to the context. Other approaches, such as random sampling, may present challenges if the user is engaged in an activity that impedes precise measurement, is not wearing the device at that moment, or if the sample does not adequately represent the range of situations experienced throughout the day. Additionally, it is well-known that the ANS undergoes 24-hour cyclic changes, known as circadian rhythms, which should also be considered. However, it is worth noting that circadian rhythms do not uniformly affect all ANS systems, nor do they present in the same way in all individuals; individual factors such as sleep and activity patterns may influence ANS rhythms, potentially requiring personalized adjustments in measurements [23, 24]. Although this methodology generates smaller volumes of data, it facilitates interpretation, processing, and storage. Moreover, it allows for a certain level of labeling, linking ANS states with a specific activity or emotion through a notification system that prompts the user to report their current state.

Another approach might involve controlled data collection, in which the subject is asked to perform specific tasks designed to induce stress or relaxation, thus allowing changes to be measured in a controlled manner. However, this approach has the drawback of depending on the user's collaboration and willingness to complete the tasks. Additionally, a habituation phenomenon may occur, where repeated exposure to stress-inducing tasks reduces their effect in subsequent applications. It should be noted that the degree of habituation may vary depending on the task, the interval between sessions, and the individual; therefore, a rigorous protocol could include task variations to minimize habituation in longitudinal studies.

This leads to the next challenge: adherence to monitoring. All monitoring

or treatment processes face adherence issues, but these may be particularly complex in patients with mental disorders, such as anxiety or depression. The design of notifications, tasks, and even the feedback provided to the patient must be carefully studied so that monitoring is not intrusive and encourages its use. Surveys and other data collected via smartphones must strike a balance that allows for the necessary information to be obtained without becoming tedious. As mentioned, these surveys can provide context to objective measurements. For example, they may include questions about recent activity, current emotional state, whether the user is indoors or outdoors, etc. Monitoring the emotional state of patients may also provide independent value to healthcare professionals and can be carried out through standardized tests, keywords, diaries, or even the use of emoticons.

Another major challenge is the complexity and heterogeneity of mental disorders. The symptoms experienced by patients may be diverse, making it necessary to investigate their relationship with available metrics and to develop new metrics that enable a deeper understanding of how these symptoms interact to shape the clinical picture.

Finally, this set of challenges is complemented by the intrinsic difficulties of using wearable devices for health monitoring. Issues such as robustness in data analysis, usability, data security, and cost must also be considered along with the previously mentioned aspects. In particular, the protection of data collected by these devices requires careful design of information flow, data encryption, and compliance with privacy regulations.

2 Obstructive sleep apnea

Obstructive Sleep Apnea (OSA) is a syndrome caused by repetitive episodes of total or partial interruption of respiratory flow during sleep due to blockades produced by intermittent relaxation of the upper airway muscles (see Figure 10). Obstructive respiratory events are the cause of sleep fragmentation, hypoxemia, hypercapnia, and increased sympathetic activity [25]. The list of symptoms can include daytime sleepiness, cognitive impair-

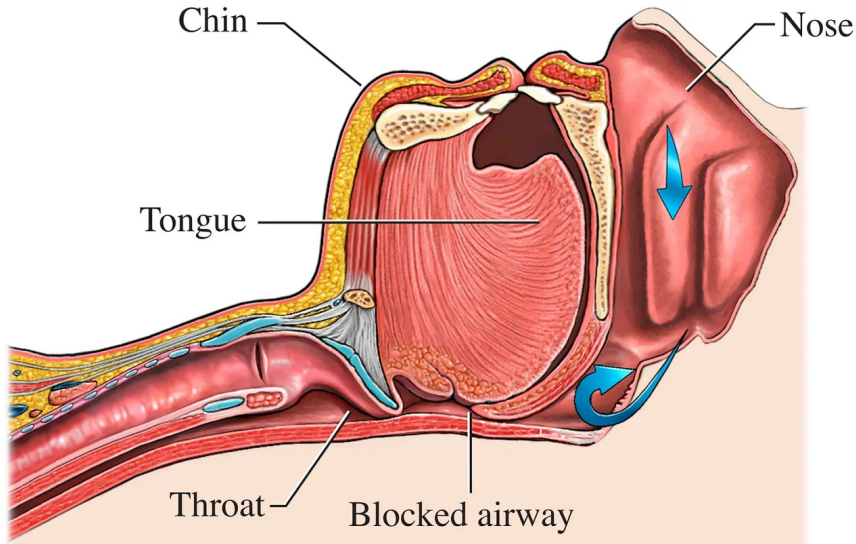


Figure 10: Airway obstruction during OSA. Extracted from [33].

ment, memory loss [26], together with comorbidities such as hypertension, cerebrovascular artery disease, coronary artery disease, congestive heart failure, and atrial fibrillation [27, 28]. Moreover, the incidence of both fatal and non-fatal cardiovascular events is substantially greater in untreated individuals with severe OSA compared to healthy individuals [29]. OSA prevalence ranges from 9% to 38% in the general adult population, being much higher in the elderly groups [30]. Furthermore, prevalence is expected to increase in the general population due to the obesity and overweight epidemic [31]. OSA underdiagnosis was estimated as 93% for women and 82% for men by Young *et al.* [32]. The rise in obesity rates combined with the widespread implementation of screening may have influenced these statistics over time.

For decades, the gold standard for diagnosis included polysomnography (PSG) performed in a clinical environment. The patient is requested to sleep in a medical center while constantly monitored, making this test uncomfortable and having some impact on natural sleep. Recently, the use of out-of-center sleep testing with limited channels was included in the diagnostic criteria for adult OSA, although it commonly underestimates the number of obstructive respiratory events per hour compared to PSG [27].

Obstructive respiratory events are usually measured by the apnea-hypopnea index (AHI). This index, which is the total count of apneas and hypopneas normalized by sleep time in hours, has been a matter of controversy since its introduction in OSA diagnosis and severity rating [34, 35].

Patients are classified as having mild, moderate, or severe OSA based on their AHI score. However, despite its widespread clinical use, the AHI may not always capture the true clinical impact of OSA, which has prompted growing concern among researchers and clinicians. One limitation of the AHI is its focus solely on event frequency without considering the duration of each respiratory event. Events that last longer can cause more significant oxygen desaturation and physiological stress compared to shorter events. This means that two patients with the same AHI may experience different levels of physiological impact depending on the duration of their apnea and hypopnea episodes. As a result, relying on AHI alone could lead to an underestimation of the disease's severity for patients with prolonged respiratory events, potentially affecting treatment decisions. Moreover, the AHI does not measure the degree of oxygen desaturation associated with each episode. In clinical practice, patients with the same AHI might have very different levels of oxygen desaturation, which can significantly impact cardiovascular health. Studies have shown that low oxygen levels during sleep are linked to higher risks of hypertension, stroke, and other cardiovascular diseases. By overlooking oxygen desaturation levels, the AHI may not fully represent the cardiovascular burden of OSA, potentially limiting its effectiveness in identifying high-risk patients who might benefit from early intervention. Additionally, the AHI does not capture daytime symptoms, such as excessive sleepiness, fatigue, or cognitive impairment, which are common in patients with OSA. These symptoms can significantly affect a patient's quality of life and have serious implications for public health, as they contribute to higher risks of motor vehicle accidents and reduced work productivity [36]. The absence of symptomatology assessment in the AHI makes it challenging to assess the full impact of OSA on a patient's daily functioning, which can lead to insufficient treatment for those who suffer primarily from these non-respiratory symptoms. Overall,

while the AHI has been useful for the initial categorization of OSA, its limitations in representing the multidimensional nature of the disease call for a more comprehensive diagnostic approach. Researchers have proposed alternative indices that incorporate event duration, oxygen desaturation, and symptomatology to provide a more complete assessment of OSA severity. Integrating these aspects could lead to improved treatment strategies and ultimately better outcomes for patients with OSA. Despite this, AHI remains the main measurement in OSA diagnosis [27].

2.1 Wearable monitoring

The early diagnosis of OSA is important, as it can cause several major health problems [37]. The underdiagnosis of OSA would be reduced by developing novel techniques for massive screenings in the general population. Among these techniques, HR dynamics assessment is appealing, as it can be applied to signals recorded at home using wearables. The variability of the heartbeat period is known to be related to sleep breathing disorders. Zwillich *et al.* [38] discovered that most apneas, excluding those without oxygen desaturations, are associated with bradycardia episodes and that bradycardias became more marked when the length of the apnea and the desaturation of oxyhemoglobin increase. In 1984, Guilleminault *et al.* [39] described the cyclic variation of the heart rate (CVHR), a bradycardia pattern during apnea, followed by abrupt tachycardia in restoration of airflow. This pattern has been an object of study, including frequency-domain analysis [40], morphology variations [41] and automatic detection [42]. Shiomi *et al.* [40] discovered an augmented very low frequency (VLF) component of heart rate (0.008-0.04 Hz) in OSA patients synchronized with episodes of absence of air exchange or hypoxemia, that occurred at a cycle length of 25-120 seconds. They also described a VLF peak during OSA episodes, likely related to the CVHR oscillation frequency, itself related to the frequency of apneas. Stein *et al.* [41] set a 20% of sleep time with CVHR as a threshold to predict $AHI \geq 15$. They proposed that, although Guilleminault identified CVHR as a valuable method for observing OSA, it was not implemented by the late

twentieth century because of technological challenges. They also pointed out back in 2003 that this technique should be included as a part of routine Holter reports.

PRV is a well-known alternative that offers a high correlation with HRV even in non-stationary situations [43]. Khandoker *et al.* [26] demonstrated that PRV could be used to distinguish OSA events from normal breathing during sleep, although several variability measures were significantly different from the HRV reference during OSA events. Analogously, Lázaro *et al.* [44] demonstrated that PRV can be used as an HRV surrogate in apnea detectors based on decreases of amplitude fluctuations of the PPG (DAP). Later, Lazazzera *et al.* [45] combined DAP, PRV, and peripheral oxygen saturation (SpO_2) for the screening of OSA in adults. In [46], Hayano *et al.* presented an automatic detection of the CVHR pattern from a PPG signal for its use in a commercial wearable watch device. This algorithm is based on the detection of every cycle on the pulse-to-pulse interval (PPI) signal. Magnúsdóttir *et al.* [47] used CVHR combined with cardiopulmonary coupling to identify sleep apnea.

2.2 Main challenges

The monitoring of sleep apnea faces several important challenges due to the complexity of the condition and the current technological limitations. Obtaining a precise and accessible diagnosis is a major challenge. Home-based monitoring devices offer convenience, but often lack the precision and detailed data for a complete diagnosis, leading to partial evaluations. Continuous noninvasive monitoring remains a major challenge in evaluating sleep apnea, and the progress in creating noninvasive sensors that provide both comfort and precise long-term information is ongoing. Although some wearable devices can track sleep, they typically have difficulty accurately identifying sleep apnea.

Wearable monitoring is challenged by the occurrence of motion artifacts, even during periods when the user is at rest. Involuntary nocturnal movements, particularly in patients with disturbed sleep, often involve the limbs and

can affect signal quality. Therefore, accurate treatment of these motion artifacts is essential to ensure reliable data processing. In addition, precise monitoring requires the observation of various elements such as breathing, airflow, oxygen saturation, and cardiac rhythm. The absence of signals such as the electroencephalogram or nasal pressure poses significant challenges in identifying arousals or conducting sleep staging.

A notable challenge in the detection of respiratory events is the variability in the annotation of events between experts [48–50]. This inconsistency reveals a significant degree of subjectivity and highlights that event labeling is highly dependent on the interpretation of the individual scorer. Factors such as differences in experience, understanding of event definitions, and even institutional guidelines can contribute to these discrepancies. This variability complicates the process of building reliable datasets, which are essential for training and validating detection algorithms. The lack of consistency in labeling makes it difficult to standardize sleep apnea assessments and creates additional obstacles for developing automated detection systems. For machine learning algorithms, in particular, training on inconsistently labeled data may reduce their performance, as the algorithms might learn to detect patterns that are more specific to the labeling style of individual scorers than to the actual physiological events.

This variability is also explained by methodological reasons. The American Academy of Sleep Medicine (AASM) establishes a standardized framework for scoring apnea and hypopnea events; however, certain aspects remain open to interpretation. For example, the AASM guidelines do not detail the exact methodology for computing airflow based on nasal pressure signals. Different institutions may employ different algorithms, such as low-pass filtering or peak detection, to obtain the envelope of this signal, leading to inconsistencies in how airflow reduction is perceived and scored in different patients. Furthermore, the AASM guidelines do not provide specific instructions on how to establish a standard baseline for airflow to use as a reference for evaluating reductions. Similarly, they lack guidance on setting a baseline for SpO₂. Without a clear method for determining this baseline, clinicians and technicians face difficulties in accurately quantifying

airflow reductions. This baseline is essential because it allows scorers to determine the percentage reduction in airflow necessary to classify events as hypopneas or apneas. Different approaches to establishing the baseline can lead to significant variations in the estimated severity of respiratory events. This further reduces the reproducibility of apnea severity assessments, complicating comparisons between studies or clinical settings. These ambiguities underscore the potential value of wearable devices that employ consistent, automated algorithms to detect and score apneic events in a standardized manner. With a carefully calibrated and transparent algorithm, devices could minimize subjectivity by automating the processing of nasal pressure signals, calculating a stable baseline, and establishing standardized thresholds for event classification. This consistency could improve the reliability of apnea severity metrics, allowing clinicians to track changes in a patient's condition with greater precision and ultimately contributing to more uniform and objective assessments in various clinical settings.

Early detection and personalized treatment present additional challenges. Adherence to the use of a wearable device as a screening tool in daily life presents a significant challenge, particularly in the context of chronic conditions such as sleep apnea. For a wearable device to be effective, it must strike a balance between functionality and comfort. The device should be minimally intrusive to avoid disrupting the user's daily activities or sleep patterns. This means that it should be lightweight, comfortable, and easy to wear, with a design that seamlessly integrates into the user's lifestyle. In addition to physical comfort, the device must offer meaningful real-time feedback to engage the user actively. Feedback mechanisms, such as personalized alerts or progress summaries, can motivate users by providing insight into their health trends and the impact of positive behavioral changes. Furthermore, long-term adherence often depends on the perceived utility of the device beyond just data collection. If the wearable provides actionable information, such as tips to improve sleep quality or reminders about healthy sleep practices, it becomes a tool for proactive health management rather than merely a diagnostic device. Such features can significantly improve user engagement and adherence over time, leading to more reliable data

collection for clinicians.

Regarding personalized treatment, sleep apnea exhibits significant variability between individuals. Monitoring systems must adapt to these differences to provide accurate and individualized diagnoses [51]. In addition, integrating monitoring with treatment is critical. For example, optimizing therapies such as Continuous Positive Airway Pressure (CPAP) requires real-time adjustments of device parameters based on current monitoring data [52]. With more comprehensive long-term data, clinicians can adjust treatment plans based on real-world evidence of the patient's sleep behavior rather than relying solely on isolated polysomnography results from a single night in a clinical setting [53, 54]. When integrated with CPAP or other therapeutic devices, wearables could enable real-time feedback loops, automatically adjusting treatment parameters in response to changes in physiological states during sleep.

Lastly, data privacy and security are crucial concerns in monitoring sleep apnea. Sleep monitoring data is sensitive and must be adequately protected to ensure patient privacy. Secure data transmission and storage are particularly important in home environments and with devices connected to the cloud, as the potential for breaches and unauthorized access remains a significant risk. These protections are especially relevant for individuals in professions where sleep apnea can be disabling, such as truck drivers, whose roles demand high levels of alertness for public safety. In addition, in such cases, false positives in the detection of sleep apnea could be especially harmful, potentially leading to unnecessary restrictions or job limitations based on inaccurate data. Advances in technology, more accurate and comfortable portable devices, and remote diagnostic systems are essential to overcome these challenges, promoting safer, more reliable, and accessible management of sleep apnea.

Bibliography

- [1] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association Publishing, 3 2022.
- [2] Jonathan Rottenberg. Cardiac vagal control in depression: A critical analysis. *Biological Psychology*, 74(2):200–211, 2 2007.
- [3] Jessica L. Hamilton and Lauren B. Alloy. Atypical reactivity of heart rate variability to stress and depression across development: Systematic review of the literature and directions for future research, 12 2016.
- [4] Carmen Schiweck, Deborah Piette, Daniel Berckmans, Stephan Claes, and Elske Vrieze. Heart rate and high frequency heart rate variability during stress as biomarker for clinical depression. A systematic review. *Psychological Medicine*, 49(2):200–211, 1 2019.
- [5] Sheldon Cohen, Denise Janicki-Deverts, and Gregory E Miller. Psychological Stress and Disease. *JAMA*, 298(14):1685–1687, 2007.
- [6] R. Castaldo, P. Melillo, U. Bracale, M. Caserta, M. Triassi, and L. Pecchia. Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis, 2015.
- [7] Allison A. Campbell and Blair E. Wisco. Respiratory sinus arrhythmia reactivity in anxiety and posttraumatic stress disorder: A review of literature, 7 2021.
- [8] Katharina Kircanski, Christian E. Waugh, M. Catalina Camacho, and Ian H. Gotlib. Aberrant parasympathetic stress responsivity in pure

- and co-occurring major depressive disorder and generalized anxiety disorder. *Journal of Psychopathology and Behavioral Assessment*, 38(1):5–19, 3 2016.
- [9] Kristen Salomon, April Clift, Mardís Karlsdóttir, and Jonathan Rottenberg. Major Depressive Disorder Is Associated With Attenuated Cardiovascular Reactivity and Impaired Recovery Among Those Free of Cardiovascular Disease. *Health Psychology*, 28(2):157–165, 3 2009.
- [10] Anna C. Phillips, Kate Hunt, Geoff Der, and Douglas Carroll. Blunted cardiac reactions to acute psychological stress predict symptoms of depression five years later: Evidence from a large community study. *Psychophysiology*, 48(1):142–148, 1 2011.
- [11] Ryan C. Brindle, Annie T. Ginty, and Sarah M. Conklin. Is the association between depression and blunted cardiovascular stress reactions mediated by perceptions of stress? *International Journal of Psychophysiology*, 90(1):66–72, 10 2013.
- [12] Aaron J. Fisher and Michelle G. Newman. Heart rate and autonomic response to stress after experimental induction of worry versus relaxation in healthy, high-worry, and generalized anxiety disorder individuals. *Biological Psychology*, 93(1):65–74, 4 2013.
- [13] David F. Tolin, Eric Lee, Hannah C. Levy, Akanksha Das, Liya Mammo, Benjamin W. Katz, and Gretchen J. Diefenbach. Psychophysiological assessment of stress reactivity and recovery in anxiety disorders. *Journal of Anxiety Disorders*, 82, 8 2021.
- [14] Spyridon Kontaxis, Eduardo Gil, Vaidotas Marozas, Jesus Lazaro, Esther Garcia, Mar Posadas-De Miguel, Sara Siddi, Maria Luisa Bernal, Jordi Aguilo, Josep Maria Haro, Concepcion De La Camara, Pablo Laguna, and Raquel Bailon. Photoplethysmographic Waveform Analysis for Autonomic Reactivity Assessment in Depression. *IEEE Transactions on Biomedical Engineering*, 68(4):1273–1281, 4 2021.
- [15] David Sheffield, Rungroj Krittayaphong, Wayne E Cascio, Kathleen C Light, Robert N Golden, Jerry B Finkel, George Gtekas, Gary G Koch,

- David S Sheps, Rungroj E Kfittayaphong Wayne, George Glekax, and Kath-leen C Light. Heart Rate Variability at Rest and During Mental Stress in Patients With Coronary Artery Disease: Differences in Patients With High and Low Depression Scores. *International Journal of Behavioral Medicine*, 5(1):31–47, 1998.
- [16] Everard W. Thornton and C. N. Hallas. Affective status following myocardial infarction can predict long-term heart rate variability and blood pressure reactivity. *British Journal of Health Psychology*, 4(3):231–245, 9 1999.
- [17] Scott C. Matthews, Richard A. Nelesen, and Joel E. Dimsdale. Depressive symptoms are associated with increased systemic vascular resistance to stress. *Psychosomatic Medicine*, 67(4):509–513, 2005.
- [18] Mandy X. Hu, Femke Lamers, Eco J.C. De Geus, and Brenda W.J.H. Penninx. Differential autonomic nervous system reactivity in depression and anxiety during stress depending on type of stressor. *Psychosomatic Medicine*, 78(5):562–572, 6 2016.
- [19] Jeffrey L Kibler and Mindy Ma. Depressive Symptoms and Cardiovascular Reactivity to Laboratory Behavioral Stress. *International Journal of Behavioral Medicine*, 11(2):81–87, 2004.
- [20] Ronald C Kessler. The Effects of Stressful Life Events on Depression. *Annual Review of Psychology*, 48:191–214, 1997.
- [21] Vaishnav Krishnan and Eric J. Nestler. The molecular neurobiology of depression. *Nature*, 455(7215):894–902, 10 2008.
- [22] Melanie Harris, Nick Glozier, Rajeev Ratnavadivel, and Ronald R. Grunstein. Obstructive sleep apnea and depression, 12 2009.
- [23] M. Wittmann, J. Dinich, M. Mellow, and T. Roenneberg. Social jetlag: Misalignment of biological and social time. *Chronobiology International*, 23(1-2):497–509, 2006.
- [24] M. Mellow, K. Spoelstra, and T. Roenneberg. The circadian cycle: Daily rhythms from behaviour to genes. *EMBO Reports*, 6(10):930–935, 2005.

- [25] Task Force of The European Society of Cardiology and the North American Society for Pacing and Electrophysiology. Heart rate variability Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3):354–381, 1996.
- [26] Ahsan H. Khandoker, Chandan K. Karmakar, and Marimuthu Palaniswami. Comparison of pulse rate variability with heart rate variability during obstructive sleep apnea. *Medical Engineering and Physics*, 33(2):204–209, 3 2011.
- [27] AASM. International classification of sleep disorders. *Darien, IL: American Academy of Sleep Medicine*, 3rd edition, 2014.
- [28] José M. Marin, Alvar Agusti, Isabel Villar, Marta Forner, David Nieto, Santiago J. Carrizo, Ferran Barbé, Eugenio Vicente, Ying Wei, F. Javier Nieto, and Sanja Jelic. Association Between Treated and Untreated Obstructive Sleep Apnea and Risk of Hypertension. *JAMA*, 307(20), 5 2012.
- [29] J Marin, S Carrizo, E Vicente, and A Agusti. Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study. *The Lancet*, 365(9464):1046–1053, 3 2005.
- [30] Chamara V. Senaratna, Jennifer L. Perret, Caroline J. Lodge, Adrian J. Lowe, Brittany E. Campbell, Melanie C. Matheson, Garun S. Hamilton, and Shyamali C. Dharmage. Prevalence of obstructive sleep apnea in the general population: A systematic review. *Sleep Medicine Reviews*, 34:70–81, 8 2017.
- [31] Paul E. Peppard, Terry Young, Jodi H. Barnet, Mari Palta, Erika W. Hagen, and Khin Mae Hla. Increased prevalence of sleep-disordered breathing in adults. *American Journal of Epidemiology*, 177(9):1006–1014, 5 2013.
- [32] Terry Young, Linda Evans, Laurel Finn, and Mari Palta. Estimation of the Clinically Diagnosed Proportion of Sleep Apnea Syndrome in Middle-aged Men and Women. *Sleep*, 20(9):705–706, 1997.

- [33] Sheetu Singh. Know Some Important Points About Obstructive Sleep Apnea, 9 2023.
- [34] Dirk A. Pevernagie, Barbara Gnidovec-Strazisar, Ludger Grote, Raphael Heinzer, Walter T. McNicholas, Thomas Penzel, Winfried Randerath, Sophia Schiza, Johan Verbraecken, and Erna S. Arnardottir. On the rise and fall of the apnea-hypopnea index: A historical review and critical appraisal. *Journal of Sleep Research*, 29(4):e13066, 8 2020.
- [35] Winfried Randerath, Claudio L. Bassetti, Maria R. Bonsignore, Ramon Farre, Luigi Ferini-Strambi, Ludger Grote, Jan Hedner, Malcolm Kohler, Miguel Angel Martinez-Garcia, Stefan Mihaicuta, Josep Montserrat, Jean Louis Pepin, Dirk Pevernagie, Fabio Pizza, Olli Polo, Renata Riha, Silke Ryan, Johan Verbraecken, and Walter T. McNicholas. Challenges and perspectives in obstructive sleep apnoea. *European Respiratory Journal*, 52(3), 2018.
- [36] S. Garbarino, O. Guglielmi, A. Sanna, G.L. Mancardi, and N. Magnavita. Risk of occupational accidents in workers with obstructive sleep apnea: Systematic review and meta-Analysis. *Sleep*, 39(6):1211–1218, 2016.
- [37] Sahar Ahmadzadeh, Jikui Luo, and Richard Wiffen. Review on Biomedical Sensors, Technologies and Algorithms for Diagnosis of Sleep Disordered Breathing: Comprehensive Survey. *IEEE Reviews in Biomedical Engineering*, 15:4–22, 2022.
- [38] C Zwillich, T Devlin, D White, N Douglas, J Weil, and R Martin. Bradycardia during sleep apnea. Characteristics and mechanism. *Journal of Clinical Investigation*, 69(6):1286–1292, 1982.
- [39] Christian Guilleminault, Roger Winkle, Stuart Connolly, Kenneth Melvin, and Ara Tilkian. CYCLICAL VARIATION OF THE HEART RATE IN SLEEP APNOEA SYNDROME: Mechanisms, and Usefulness of 24 h Electrocardiography as a Screening Technique. *The Lancet*, 323(8369):126–131, 1984.
- [40] Ttoshiaki Shiomi, Tchristian Guilleminault, Ryujiro Sasanabe, Izumi Hirota, Masato Maekawa, and Tadashi Kobayashi. Augmented Very

- Low Frequency Component of Heart Rate Variability During Obstructive Sleep Apnea. *Sleep*, 19(5):370–377, 1996.
- [41] Phyllis K Stein, Stephen P Duntley, Peter P Domitrovich, Pallavi Nishith, and Robert M Carney. A Simple Method to Identify Sleep Apnea Using Holter Recordings Identifying Sleep Apnea Using Holter Recordings. Introduction: Despite its severe cardiovascular. *J Cardiovasc Electrophysiol*, 14:467–473, 2003.
- [42] Junichiro Hayano, Eiichi Watanabe, Yuji Saito, Fumihiko Sasaki, Keisaku Fujimoto, Tetsuo Nomiyama, Kiyohiro Kawai, Itsuo Kodama, and Hiroki Sakakibara. Screening for obstructive sleep apnea by cyclic variation of heart rate. *Circulation: Arrhythmia and Electrophysiology*, 4(1):64–72, 2 2011.
- [43] E. Gil, M. Orini, R. Bailón, J. M. Vergara, L. Mainardi, and P. Laguna. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. *Physiological Measurement*, 31(9):1271–1290, 2010.
- [44] Jesús Lázaro, Eduardo Gil, José María Vergara, and Pablo Laguna. Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children. *IEEE Journal of Biomedical and Health Informatics*, 18(1):240–246, 2014.
- [45] Remo Lazazzera, Margot Deviaene, Carolina Varon, Bertien Buyse, Dries Testelmans, Pablo Laguna, Eduardo Gil, and Guy Carrault. Detection and Classification of Sleep Apnea and Hypopnea Using PPG and SpO2 Signals. *IEEE Transactions on Biomedical Engineering*, 68(5):1496–1506, 5 2021.
- [46] Junichiro Hayano, Hiroaki Yamamoto, Izumi Nonaka, Makoto Komazawa, Kenichi Itao, Norihiro Ueda, Haruhito Tanaka, and Emi Yuda. Quantitative detection of sleep apnea with wearable watch device. *PLoS ONE*, 15(11 November), 11 2020.

- [47] Solveig Magnúsdóttir and Hugli Hilmisson. Ambulatory screening tool for sleep apnea: analyzing a single-lead electrocardiogram signal (ECG). *Sleep and Breathing*, 22(2):421–429, 5 2018.
- [48] J.B. Stephansen, A.N. Olesen, M. Olsen, A. Ambati, E.B. Leary, H.E. Moore, O. Carrillo, L. Lin, F. Han, H. Yan, P. Jennum, and E. Mignot. Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. *Nature Communications*, 9(1), 2018.
- [49] A. Guillot, F. Sauvet, E.H. During, and V. Thorey. DREAM Open Datasets: Multi-Scored Sleep Datasets to Compare Human and Automated Sleep Staging. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(9):1955–1965, 2020.
- [50] R.S. Rosenberg and S. Van Hout. The American Academy of Sleep Medicine inter-scorer reliability program: Respiratory events. *Journal of Clinical Sleep Medicine*, 10(4):447–454, 2014.
- [51] B.A. Edwards, S. Landry, S.A. Joosten, and G.S. Hamilton. Personalized Medicine for Obstructive Sleep Apnea Therapies: Are We There Yet? *Sleep Medicine Clinics*, 11(3):299–311, 2016.
- [52] S. Sharma, S. Wali, Z. Pouliot, M. Peters, H. Neufeld, and M. Kryger. Treatment of obstructive sleep apnea with a self-titrating continuous positive airway pressure (CPAP) system. *Sleep*, 19(6):497–501, 1996.
- [53] C.A. Kushida, A. Chediak, R.B. Berry, L.K. Brown, D. Gozal, C. Iber, S. Parthasarathy, S.F. Quan, and J.A. Rowley. Clinical guidelines for the manual titration of positive airway pressure in patients with obstructive sleep apnea. *Journal of Clinical Sleep Medicine*, 4(2):157–171, 2008.
- [54] C. Schöbel, S. Knorre, M. Glos, C. Garcia, I. Fietze, and T. Penzel. Improved follow-up by peripheral arterial tonometry in cpap-treated patients with obstructive sleep apnea and persistent excessive daytime sleepiness. *Sleep and Breathing*, 22(4):1153–1160, 2018.

General methodology for robust ANS monitorization

1 Methodological framework

Robust monitoring of ANS through HRV using wearable devices requires an adaptation of the entire processing chain. Signals from these devices often have a lower signal-to-noise ratio compared to their analogous clinical devices, as they are exposed to continuous changes in the environment and movement artifacts. Devices like smartwatches are designed to be worn throughout the day, both at rest and during movement, walking, running, or performing daily tasks. For a wearable device to be adopted by users, it must ensure a certain level of comfort, which is not as necessary in clinical devices. This implies less tight sensors and less intrusive measurement points. Smartwatches are again a good example of this. Compared to a pulse oximeter placed on the finger, a point with a high capillary density, using a clip that provides a firm hold, the PPG of a smartwatch is obtained on the wrist, with much lower capillary density, and usually a looser fit. Additionally, it is a reflective PPG, in contrast to the transmissive configuration of pulse oximeters. Although both signals are PPG and measure changes in blood volume at nearby peripheral points, the processing has to be necessarily different.

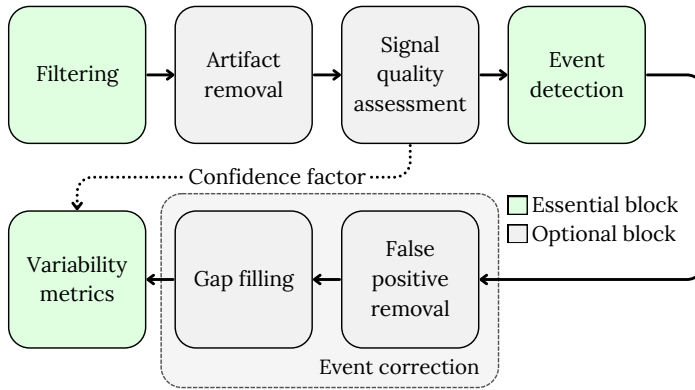


Figure 11: Processing chain for HRV metrics. Essential blocks in green, optional in grey. Within the framework of comprehensive ANS evaluation employing wearable technology, the components that are optional in controlled environments become essential.

The simplest processing chain (see Figure 11) of an ECG/PPG signal to obtain HRV/PRV consists of beat/pulse detection and metric extraction. These metrics can be as simple as the mean or standard deviation of the detection series. Typically, a filtering stage is included before detection to remove high-frequency noise and baseline wander. This chain may be sufficient in a clinical setting, with the patient at rest and analyzing metrics such as the ones mentioned, which are statistically very robust.

It is necessary to acknowledge the importance of this processing block, particularly because most event detection algorithms are typically applied to signals that lack a baseline. The application of high-frequency denoising is considered trivial, yet highly advantageous, to the extent that omitting it would be a significant oversight. In reality, the concept of denoising is more complex than this, since a filter with these characteristics removes only noise from frequencies outside the band where physiological information is expected to be found. Therefore, it would be appropriate to consider a separate denoising processing block following the filtering block to remove the noise in the targeted band. Although this thesis does not address this block, it is worth noting that there are sophisticated denoising algorithms available. The application of these algorithms may be optional, although

their use can be highly beneficial, especially in the context of wearable devices. For example, techniques that involve deep learning and adaptive filtering have shown promising results in improving signal quality while preserving important characteristics [1–3].

As the setting becomes less controlled and more complex metrics need to be analyzed, it is necessary to include additional processing blocks. Two main sets of blocks can be distinguished: those introduced before beat/pulse detection and those introduced afterward.

Accurate detection faces challenges such as movement artifacts, which occur briefly and entirely obscure the signal, as well as low-quality segments. The latter can be defined as segments with a poor signal-to-noise ratio that last longer than artifacts. This usually occurs due to poor sensor attachment, displacement from the optimal measurement point, or moderate to light movements where the signal is not completely lost. Although high-energy artifacts may seem the most harmful, they are actually the easiest to detect and therefore to eliminate. On the other hand, borderline segments can be overlooked and produce false positives in detection. As will be seen in later chapters, this type of error is the most harmful to robust monitoring. To address these issues, it is recommended to add two blocks: one for artifact removal and the other for signal quality assessment.

Both blocks are intended to select the signal segments where beat/pulse detection is performed, discarding the rest. The artifact removal block should completely discard the affected segments while allowing non-corrupted segments for further processing. This is not always easy to achieve and becomes more challenging as the energy of the artifacts more closely resembles that of the signal and as their duration decreases. In the literature, there are also methods that aim to clean corrupted segments using information before and after the event [4]. This approach is not recommended for obtaining robust HRV metrics, as it can lead to false detections or significantly displaced detections from the actual beat/pulse (see Figure 12). Instead, the goal should always be to obtain a detection series that is as reliable as possible, without the presence of false positives, even if this results in an increase in false negatives.

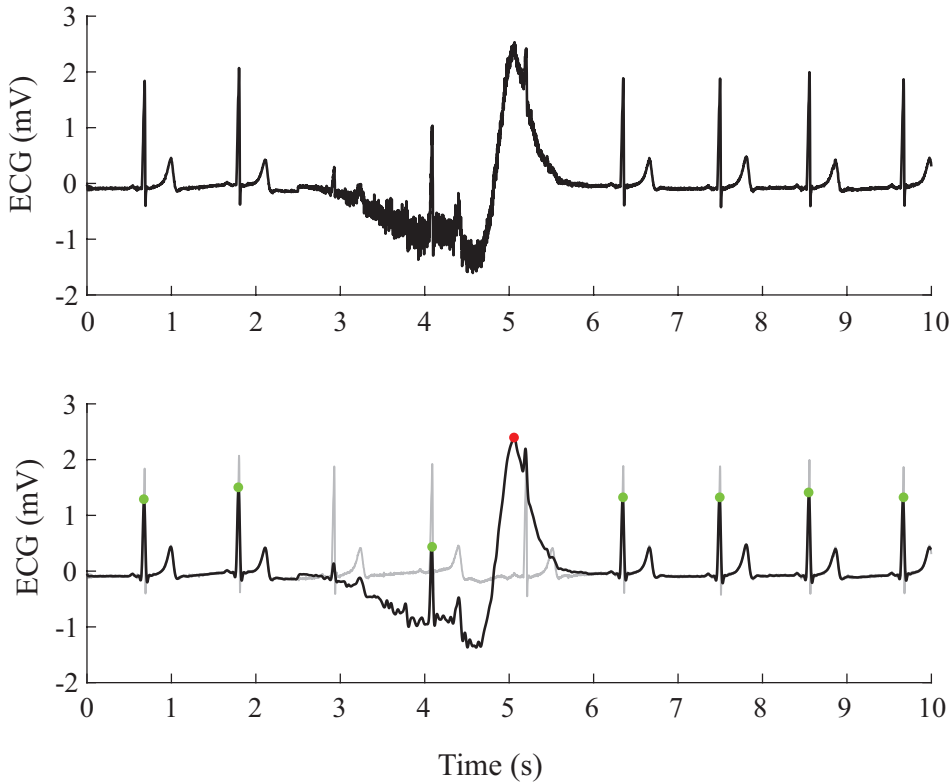


Figure 12: Erroneous detection in an ECG affected by artifacts. *Top:* Corrupted ECG. The artifact has been simulated in order to compare the actual beats with the automatically detected ones. *Bottom:* The filtered ECG is shown in black, while the clean ECG appears in gray. Green dots indicate correct detections, and red dots represent incorrect ones. The third heartbeat is missed due to a significant reduction in the QRS amplitude. The fifth heartbeat is misaligned because the artifact mimics a short-duration peak, masking the QRS complex.

On the other hand, the signal quality assessment block should provide a measure of how reliable the detections in a given segment can be, allowing the option to either discard segments or continue processing them, while associating a confidence level with the metrics. Again, the decision becomes more difficult as the signal-to-noise ratio approaches the borderline. In most cases, it is preferable to use a criterion to discard borderline segments, as they can produce metrics that seem reliable, leading to incorrect conclusions. Furthermore, trying to utilize the maximum

number of segments undermines one of the main advantages of wearable devices, namely the large volume of data they provide.

Consider the case of a sleep apnea screening tool. If the user wears a smartwatch throughout the night, it is likely they will do so on subsequent nights. Therefore, if during the first night the user constantly moves or wears the watch strap loosely, the quality assessment block should mark the signals as unusable, opting to wait for the next night to take measurements. Similarly, analyses can be performed over several consecutive nights before issuing a verdict. Leveraging the volume of measurements and redundancy should be a key principle in applications involving continuous-use wearables.

After these two blocks, a beat/pulse detector should process the clean segments. This detector produces a series of events where false positives and false negatives may occur. These can cause distortions in variability metrics, potentially leading to different effects depending on the specific metric. These effects, as well as methods for error correction, are the focus of the study presented in Chapter 6.

2 Artifact removal in PPG signals

Motion artifact removal is an absolutely necessary task to obtain robust metrics in wearable environments. Devices are continuously exposed to user movements, regardless of the type of sensor used. From the lightest activity to intense exercise, movements can appear in the signal as segments, usually of high energy, where the waveform is completely distorted or masked. Typically, artifacts are understood as segments of limited duration associated with a specific movement, although it is possible that a series of movements, such as those performed while running or in other sports disciplines, may produce artifacts over longer intervals. Artifacts should be delimited to exclude those portions from the subsequent analysis. In this thesis, this task is defined as separate from signal quality detection, explained below, where a qualitative assessment is made on the overall signal quality. In the first case, the aim is to remove corrupted segments from the signal. In the

second, evaluate whether the rest of the signal is of sufficient quality for robust ANS analysis.

Provided a zero-mean signal, $x[n]$, one of the most elementary algorithms for delimiting artifacts is the moving median threshold. Using a sliding window strategy, the moving median $\tilde{x}[n]$ is calculated within each N -sample window centered on the sample n .

$$\tilde{x}[n] = \text{med}\{x[i]\}; \forall i \in [n - N/2 + 1, n + N/2]. \quad (4.1)$$

$\tilde{x}[n]$ is then summed with a constant value K_1 higher than 0 and used as a threshold, $\tilde{x}^{th}[n]$, for each signal sample:

$$\tilde{x}^{th}[n] = \tilde{x}[n] + K_1; K_1 \in (0, \infty). \quad (4.2)$$

In the case that any sample within the window exceeds the threshold ($\exists i \in [n - N/2 + 1, n + N/2] \mid x[i] > \tilde{x}^{th}[i]$), the entire window is discarded.

In real-time applications, it is necessary to introduce a delay of $N/2$ samples to account for future values following the current sample n . This approach ensures higher accuracy by incorporating information from both past and future samples. Conversely, using a non-delayed method, which relies solely on the $N - 1$ samples preceding n , would be less precise as it depends on values further from the target sample and ignores the information provided by subsequent samples.

This method can be very effective in stationary conditions, where the signal power does not change over time, since the median, if the window is large enough, is practically unaffected by power changes and artifacts. Alternative versions of this method use the moving standard deviation in place of the moving median [5]. The moving standard deviation $x_\sigma[n]$ is computed with an unbiased estimator as

$$x_\sigma[n] = \sqrt{\frac{\sum_i (x[i] - \tilde{x}[n])^2}{N - 1}}; \forall i \in [n - N/2 + 1, n + N/2], \quad (4.3)$$

where $\bar{x}[n]$ is the mean of the signal within the window, estimated as

$$\bar{x}[n] = \frac{1}{N} \sum_i x[i]; \forall i \in [n - N/2 + 1, n + N/2]. \quad (4.4)$$

This allows the threshold to adapt to power changes. Given the condition that $x[n]$ has zero mean, the standard deviation increases with increasing power and vice versa. A proportional threshold might be more suitable in this scenario, as it allows for a less restrictive threshold at higher power levels and a more stringent one at lower powers, thereby reducing the likelihood of false positives and negatives, respectively. The threshold, $x_\sigma^{th}[n]$, should be calculated as

$$x_\sigma^{th}[n] = K_2 \cdot x_\sigma[n]; K_2 \in (1, \infty) \quad (4.5)$$

Again, in the event that any sample within the window exceeds the threshold ($\exists i \in [n - N/2 + 1, n + N/2] \mid x[i] > x_\sigma^{th}[n]$), the entire window is discarded. The moving median absolute deviation, $x_{MAD}[n]$, can be used instead of the standard deviation, obtaining similar results. The $x_{MAD}[n]$ is computed as

$$x_{MAD}[n] = \text{med}\{|x[i] - \bar{x}[n]|\}; \forall i \in [n - N/2 + 1, n + N/2] \quad (4.6)$$

and the threshold as

$$x_{MAD}^{th}[n] = K_3 \cdot x_{MAD}[n]; K_3 \in (1, \infty). \quad (4.7)$$

It should be noted that $x_{MAD}[n]$ and $x_\sigma[n]$ are proportional if $x[n]$ is normally distributed. Figure 13 illustrates an example of the three thresholds applied to an identical segment. Observe that $\tilde{x}^{th}[n]$ is flattened relative to the others due to a non-proportional threshold.

These algorithms are very easy to implement and are highly effective against short-duration, high-energy artifacts, but are much less effective against artifacts that produce subtle distortions. They are also useful for identifying

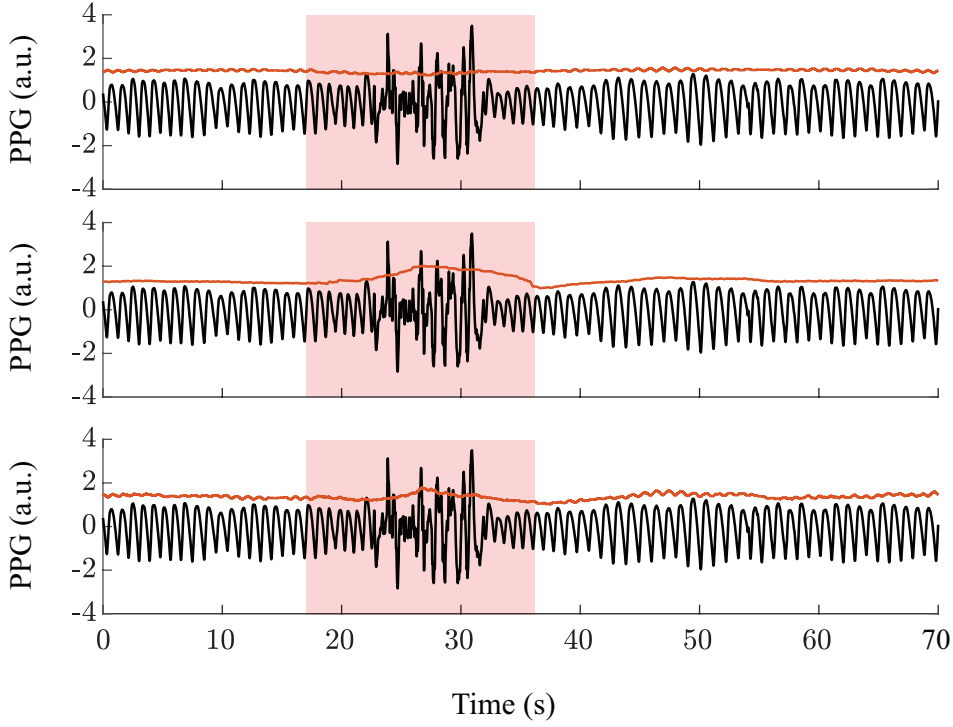


Figure 13: Artifact identification using sliding windows of 10 seconds. ($N = 2560$ at 256 Hz sampling rate). PPG in black, thresholds in red. In the event that any sample of the PPG within the 10-second interval surpasses the threshold, the entire interval is rejected (red shade). *First:* $\tilde{x}^{th}[n]$ threshold with $K_1 = 1.3$. *Second:* $x_\sigma^{th}[n]$ threshold with $K_2 = 1.5$. *Third:* $x_{MAD}^{th}[n]$ threshold with $K_3 = 2$.

outliers in detection sequences, as is covered in Section 5. The specified thresholds may be used separately or in combination. Furthermore, the text has focused solely on upper thresholds, but it is also possible to define lower thresholds in a similar manner.

2.1 Hjorth parameters

In order to remove artifacts in PPG signals, a modified version of $\tilde{x}^{th}[n]$ can be applied, focusing on parameters derived from the signal instead of the signal itself. Hjorth parameters were originally described as a method for the qualitative description of encephalographic signals [6]. Bo Hjorth

described these parameters in 1970, naming them Activity (\mathcal{H}_0), Mobility (\mathcal{H}_1), and Complexity (\mathcal{H}_2) [6]. The artifact detector used in this thesis sets upper and lower limits for the Hjorth parameters, as proposed in [6, 7], calculated from signal data within sliding windows with a certain overlap. The signal, window size, and overlap are specified in each study in the following chapters. Provided a zero-mean signal, $x[n]$, the Hjorth parameters are defined as:

$$\begin{aligned} \text{Activity : } \mathcal{H}_0[m] &= \bar{w}_0[m] \\ \text{Mobility : } \mathcal{H}_1[m] &= \sqrt{\frac{\bar{w}_2[m]}{\bar{w}_0[m]}} \\ \text{Complexity : } \mathcal{H}_2[m] &= \sqrt{\frac{\bar{w}_4[m]}{\bar{w}_2[m]} - \frac{\bar{w}_2[m]}{\bar{w}_0[m]}}, \end{aligned} \quad (4.8)$$

where \bar{w}_i represents the i -th order spectral moment. \bar{w}_i is estimated using the temporal expression of the moments in the m -th window of N samples:

$$\hat{\bar{w}}_i[m] \approx \frac{2\pi}{N} \sum_{n=(m-1)N+1}^{mN} (x^{i/2}[n])^2. \quad (4.9)$$

The thresholds are calculated as in Equation 4.2, using $\mathcal{H}_0[m]$, $\mathcal{H}_1[m]$ and $\mathcal{H}_2[m]$ extracted from the signal $x[n]$. In this scenario, the values are not assigned to each individual sample n ; instead, they are assigned to each window m . Consequently, the task is not to identify whether a sample within the window exceeds the threshold. Instead, it involves verifying that all window parameters $\mathcal{H}_0[m]$, $\mathcal{H}_1[m]$ and $\mathcal{H}_2[m]$ are within the permissible range. If any parameter fails to satisfy this criterion, the window is discarded, consistent with the methods explained previously.

Hjorth parameters are related to intelligible characteristics of the signal. Activity is a measure of the variance or average power of the signal. In the context of artifact detection, it allows to set a maximum threshold to detect energy peaks, which are typical of pronounced artifacts (see Figure 14).

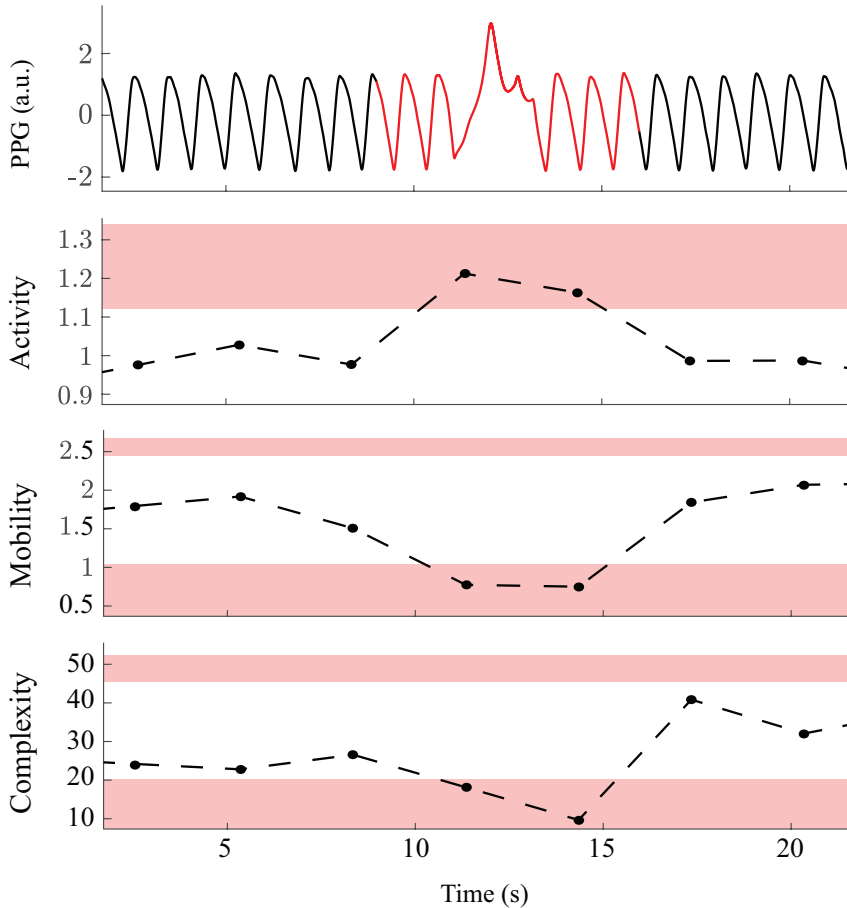


Figure 14: Artifact identification using Hjorth parameters. *First:* PPG shown in black, with automatically identified artifacts highlighted in red. Since the Hjorth parameters are calculated over a window, some uncorrupted signal segments are also eliminated. *Second, Third and Fourth:* Dashed lines with dots represent the Hjorth parameters. Note that they are sampled at each dot, representing the window center, contrary to the signal. Unallowed regions are shaded with red.

Mobility is related to the mean frequency of the signal. Signals with a low mean frequency have low mobility, and vice versa. Given that the focus is on cardiac signals, these signals are constrained to specific frequency ranges. Consequently, permissible ranges are established, beyond which the oscillations are regarded as being generated by artifacts. Complexity is related to the bandwidth of the signal spectrum. Signals that are more

similar to a sinusoid have a lower complexity than those composed of many frequencies. The presence of artifacts may increase or decrease the complexity of the signals, so upper and lower thresholds are set.

3 Signal quality assessment

In this thesis, it was not deemed necessary to include a specific signal quality assessment method for the studies presented. However, a brief review of some commonly employed methods is warranted. The necessity of such an approach is largely dependent on the application context. Generally, it can be asserted that the need for a quality assessment method increases in scenarios involving continuous monitoring, particularly in variable environments where user awareness is minimal. For example, this block may be less critical in monitoring scenarios where the user is required to perform a task for a predetermined period while being recorded. This is the case in the study presented in Chapter 7, where the user's active interest in maintaining a position that ensures a clear signal is relied upon. In these situations, it is beneficial to provide feedback to the user, such as visualizing the signal, enabling them to adjust their position or stay still to optimize signal quality. In contrast, in the context of a device like a Holter monitor or a smartwatch, which is worn throughout the day under varying conditions, the role of this block becomes more crucial. Although the user is aware of the use of the device, it is inevitable, and indeed desirable from a practical standpoint, that they do not constantly focus on it, allowing them to engage in normal daily activities. Consequently, there are periods where signal quality deteriorates significantly. Identifying these periods is essential to ensure robust monitoring.

The methodology in the literature is varied and most studies recommend the use of a quality index in some form [8–11]. These indices provide a quantitative measure of the quality evaluated, incorporating specific criteria. Such criteria may include factors ranging from similarity to a signal template to a combination of certain statistical qualities, such as skewness, peakedness,

and kurtosis, among others. Alternatively, some authors employ machine learning algorithms, such as neural networks, in their approaches [12, 13]. Despite differing methods, all share the common objective of determining whether a signal should be discarded or allowed to proceed through the processing chain. Some studies suggest specific processing algorithms based on the quality of each signal [14]. Another strategy could consist of continuing with the processing chain in any case, assigning the final metrics an uncertainty index inversely proportional to the signal quality.

4 Event detection

Observing ANS activity via HRV or PRV entails examining the fluctuation between consecutive heartbeats. Consequently, the first task is to identify the timing of each beat or pulse, based on whether the ECG or PPG signal is under review. Although the techniques employed for this purpose differ significantly, the underlying principle remains consistent: the identification of waveforms that correspond to physiological phenomena. For clarity, the term *event detection* is used to refer to this process in both ECG and PPG signals.

4.1 Event detection in ECG signals

While a full ECG provides extensive details about the heart, HRV techniques focus solely on identifying QRS complexes. These are the most easily identifiable waves due to their significant energy concentrated over a brief interval. Consequently, they serve as a dependable reference for assessing the heart's periodic activity and its variability.

A delineator based on the wavelet transform has been used to detect events in ECG signals [15]. The wavelet transform is used to decompose the ECG signal into various scales, allowing the identification of specific temporal characteristics of the ECG waves. A notable advantage of the wavelet transform lies in its ability to analyze the signal at multiple resolutions,

which is particularly useful for detecting the distinct components of the ECG, each of which has a unique frequency content. The initial detection focuses on the QRS complex, the most prominent component of the ECG. The method involves identifying the ridges of maxima in the wavelet transform that exceed predefined thresholds. To confirm the presence of a QRS complex, a zero-crossing in the wavelet transform is sought. This process is repeated across multiple scales to ensure robust detection even in the presence of noise or artifacts. Upon detecting the QRS complex, the algorithm advances to the delineation of the individual waves (P, Q, R, S, T) by identifying the peaks, as well as the onset and offset of these waves. This delineation utilizes information derived from the maxima, minima, and zero crossings of the wavelet transform across different scales. The multiscale approach inherent in the wavelet transform is central to this method, as it allows the ECG signal to be processed with varying levels of detail, thus achieving a more precise and robust delineation.

Recent advances in wavelet-based ECG analysis continue to emphasize the importance of multiscale methods. Studies have further refined these techniques, integrating machine learning to improve the accuracy of wave detection and classification. For example, deep learning models combined with wavelet transform are increasingly being used to enhance the precision of QRS complex detection, even in challenging conditions with high noise levels [16]. On the other hand, there is ongoing research into deep learning techniques that do not rely on wavelets or any kind of feature engineering, showing excellent results [17]. These approaches make use of convolutional neural networks (CNN) [18, 19], CNN combined with long short-term memory (LSTM) models [20], bilateral LSTM [21], transformers [22, 23], and the promising U-net architecture [24].

4.2 Event detection in PPG signals

The PPG waveform does not present a distinct characteristic corresponding to ventricular contraction, as observed in the QRS complex of the ECG. Notably, the PPG lacks any representation of the electrical activity of the

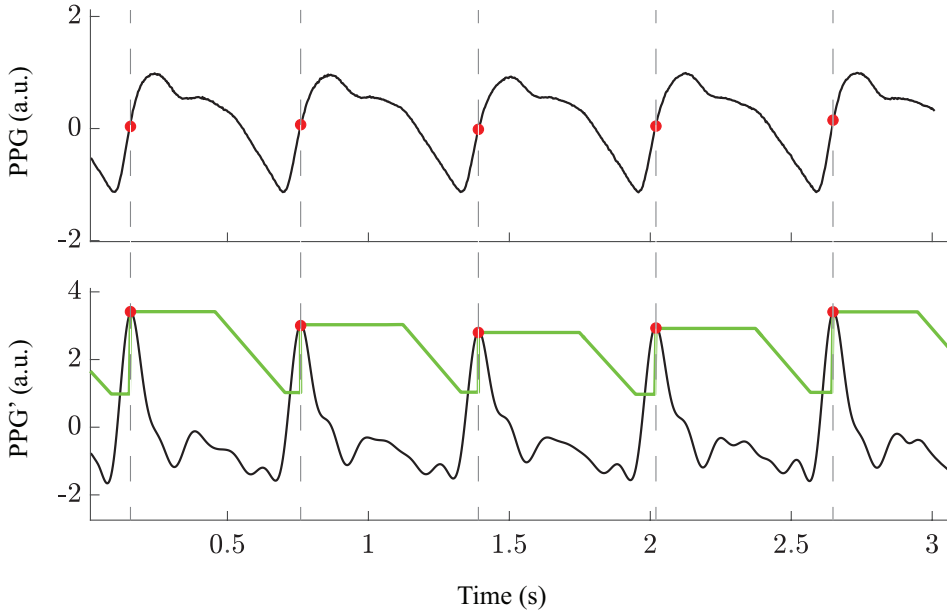


Figure 15: PPG pulse detection. *Top:* PPG signal in black with event detections in red. *Bottom:* PPG derivative in black with event detections in red. The green line shows the adaptive threshold, which is divided into three distinct phases. Following each peak detection, the threshold stays steady for a refractory period. Next, it linearly declines to a minimum value, set as a fraction of the previous maximum. Lastly, the threshold remains fixed until the subsequent detection.

heart. Instead, the information it provides is only indirectly related to cardiac function, as it reflects the pulsatile nature of blood flow in the peripheral microvasculature, primarily arterioles. This pulsatility produces a rounded waveform, in which distinguishing fiducial points becomes increasingly challenging, particularly in the presence of noise. Consequently, achieving robust monitoring is considerably more difficult in devices that record PPG, compared to those that record ECG.

The pulse detection algorithm used in this study is based on the method described in [25]. This algorithm enhances the abrupt upslope of the signal by applying a filtering transformation. Specifically, a linear-phase finite impulse response low-pass differentiation filter is utilized, which is designed using a least-squares technique. After obtaining the derivative of the PPG

signal, an adaptive thresholding process is implemented to identify the maxima.

The adaptive threshold comprises three phases (see Figure 15). Initially, following each detection, the threshold is kept constant for a refractory period of 300 ms. It then declines in a linear fashion to reach a minimum, specified as a fraction (30%) of the preceding peak value. Finally, the threshold remains constant until the subsequent detection. This technique was originally developed to manage the amplitude fluctuations commonly observed during apneic episodes, where a fixed threshold would be inadequate. However, it has also shown strong performance under other conditions [26, 27].

5 Event correction

Interval series can be represented with the interval function $d_{\text{IF}}(t_k)$, defined by

$$d_{\text{IF}}(t) = \sum_{k=1}^K (t_k - t_{k-1}) \delta(t - t_k). \quad (4.10)$$

This function is defined continuously, with zero values for all t other than t_k , such that each event occurring at time t_k is represented by a unit impulse function $\delta(t - t_k)$ scaled by the length of the preceding interval [28, 29].

Missing data detection (false negatives) is usually based on detecting physiologically abnormal increases in the interval series that suggest that at least one heartbeat is missing. Missing beats produce outliers in $d_{\text{IF}}(t_k)$ at each t_k corresponding to events after a gap. A moving median threshold is used as an outlier detection rule. First, $d_{\text{IF}}(t_k)$ is filtered with a median filter of order $2L$ to produce an expected interbeat interval $d_{\text{EIBI}}(t_k)$ for each event t_k [30]:

$$d_{\text{EIBI}}(t_k) = \text{med}\{d_{\text{IF}}(t_i)\}; \forall i \in [k - L + 1, k + L]. \quad (4.11)$$

The interval at t_k is marked as a false negative if the equation

$$d_{\text{IF}}(t_k) > (\alpha \times d_{\text{EIBI}}(t_k)) \quad (4.12)$$

is satisfied, *i.e.*, if the interval is longer than α times the expected interval, being $\alpha \in [1, \infty)$. Conversely, false positive produces an abrupt shortening of the interval as a result of the introduction of a spurious beat between two actual beats. The interval at t_k is considered a false positive if the equation

$$d_{\text{IF}}(t_k) < (\beta \times d_{\text{EIBI}}(t_k)). \quad (4.13)$$

is satisfied, *i.e.*, if the interval is shorter than β times the expected interval, being $\beta \in (0, 1)$.

Eliminating false positives is crucial, as they are often more harmful than false negatives, resulting in the loss of a reliable reference. Proper design of the prior blocks, namely artifact removal and signal quality assessment, is necessary to minimize their presence during this phase. Detecting false positives, as shown in Figure 16, proves particularly difficult when false negatives are present. The act of setting thresholds involves choosing between excluding every potential anomaly or adopting a more lenient threshold in order to preserve true positives. Nonetheless, the balance between eliminating true positives and permitting false positives should lean towards reducing false positives. Once identified, false positives should be eliminated from the t_k series, and $d_{\text{IF}}(t_k)$ recalculated.

In contrast, managing false positives is relatively straightforward, as is detailed in Chapter 6, particularly when starting with a series of reliable detections, *i.e.*, free of false positives. Because it is relatively simple to rectify a false negative, implementing a stringent threshold for false positives—which may occasionally exclude true positives—is not overly detrimental, as these instances can be addressed similarly to false negatives. This should be considered in line with the principle of "garbage in, garbage out." If one aims to perform an effective event correction, it must be acknowledged that even the best algorithm cannot correct a series of events riddled with errors.

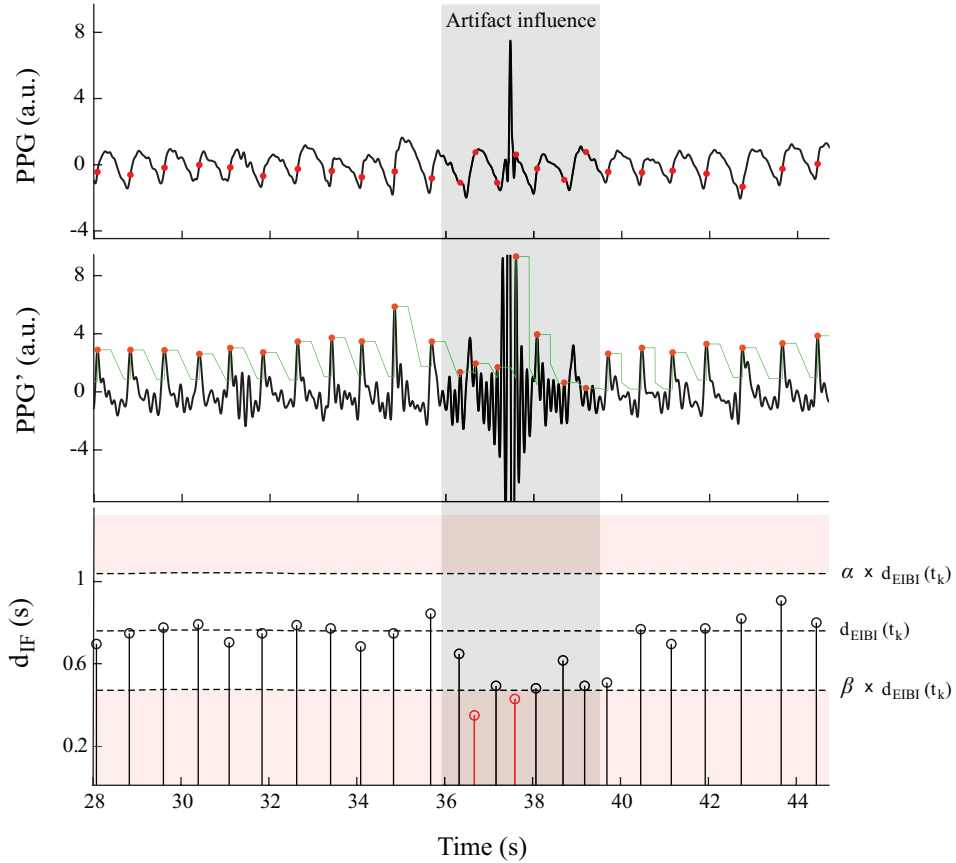


Figure 16: Outlier detection in the event series. Events located within the gray shaded area are incorrectly identified due to the artifact influence. *Top:* PPG signal. Detected pulses are marked with red dots. *Center:* PPG derivative. *Bottom:* Outlier detection in the interval function. Red shaded areas delimits values outside the allowed range. Intervals in red stems have been identified as outliers; however, some are still undetected.

5.1 Ectopic beats

Detection errors are not the only anomalies requiring correction in the event series. The presence of ectopic beats, which originate outside the SA node, introduces outliers into the data that, if not eliminated, adversely affect the metrics.

In controlled environments, such as hospitals equipped with multi-lead ECG systems, where detection errors are minimal, ectopic beats can be addressed using specialized methodologies, as in [31]. In contrast, within the context of automatic detection in variable environments, examined in this thesis, the presence of ectopic beats becomes indistinguishable from detection errors. Consequently, it is recommended to avoid employing methodologies that do not account for this overlap.

The methods outlined in this section are specifically designed to identify outliers within the series. Thus, they are equally effective for detecting ectopic beats. The subsequent treatment of these anomalies follows the same principles, both because the primary goal is to eliminate or replace outliers in line with statistical consistency, and because distinguishing between ectopic beats and detection errors is impractical in this context.

6 Variability metrics

6.1 Time domain

Methods in the time domain are the most basic and straightforward means to evaluate HRV. They involve simple statistical analyses conducted directly on the R-R interval series or the inter-beat interval (IBI) series. Calculating variability metrics from the PPG is analogous, utilizing the interval between consecutive pulses in place of the R-R interval. Another series exists, known as the normal-to-normal interval series (N-N series), which denotes the time between QRS complexes resulting from sinus node depolarization, *i.e.*, the R-R intervals after excluding ectopic heartbeats. To simplify the notation, the interval function $d_{\text{IF}}(t_k)$ is used, assuming that both ectopics and detection errors have been eliminated.

The mean heart rate (MHR), determined as the mean of the inverses of the R-R intervals, is a key indicator because it reflects the level of sinus node activation and remains a reliable metric even for segments of less than a minute [32]. It is usually measured in beats per minute (bpm). Given

$d_{\text{IF}}(t_k)$ with K events, the mean is calculated using an unbiased estimator as:

$$\overline{d_{\text{IF}}} = \frac{1}{K} \sum_{k=1}^K d_{\text{IF}}(t_k). \quad (4.14)$$

When $\overline{d_{\text{IF}}}$ is measured in seconds, the MHR, expressed in bpm, is calculated as:

$$\text{MHR} = \frac{60}{\overline{d_{\text{IF}}}}. \quad (4.15)$$

Another straightforward metric to calculate is the standard deviation of the normal-to-normal interval (SDNN). This metric corresponds to the square root of the variance and thus serves as a measure of the total variability power. SDNN, expressed in milliseconds, is calculated as:

$$\text{SDNN} = 1000 \cdot \sqrt{\frac{\sum_{k=1}^K (d_{\text{IF}}(t_k) - \overline{d_{\text{IF}}})^2}{K - 1}}, \quad (4.16)$$

SDNN, as a measure of total variability, does not differentiate which division of the autonomic nervous system influences this variation. Metrics that provide a better distinction of variability components include the root mean square of successive differences (RMSSD). This metric is computed from the differences between successive intervals rather than from the R-R intervals. As it gauges the magnitude of interval changes, it indicates the variability rate. Thus, RMSSD is intimately linked to and shows a strong correlation with the HF component of the spectrum. The RMSSD is computed in milliseconds as:

$$\text{RMSSD} = 1000 \cdot \sqrt{\frac{\sum_{k=2}^K (d_{\text{IF}}(t_k) - d_{\text{IF}}(t_{k-1}))^2}{K - 1}}. \quad (4.17)$$

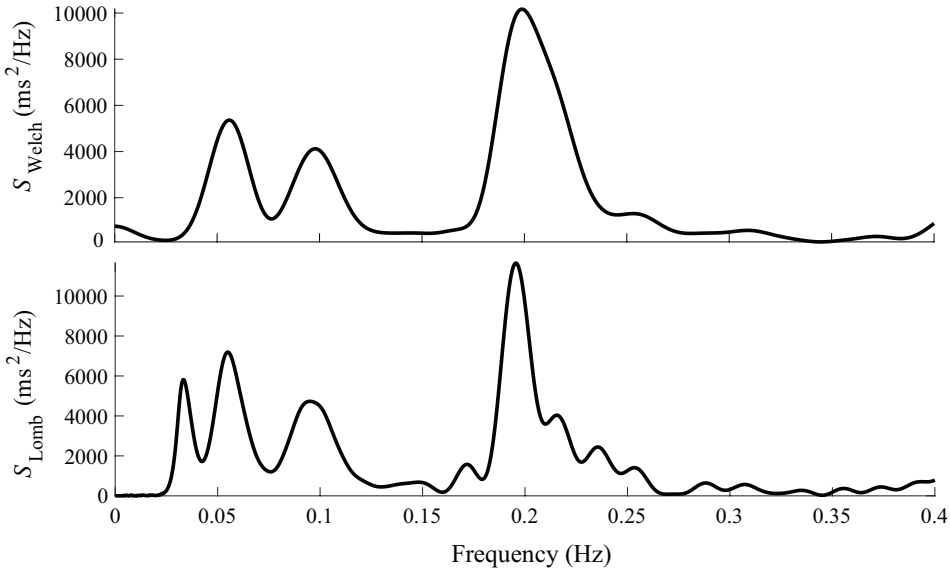


Figure 17: PSD estimates from the same record using different methods. The estimates exhibit substantial differences. *Top*: Fourier-based spectral estimation. *Bottom*: Averaged Lomb-Scargle periodograms.

6.2 Frequency domain (I). Spectral estimation

As mentioned in Chapter 2, frequency analysis allows the distinction of neural control elements within HRV. In this thesis, power spectral density (PSD) estimation has been performed using both the fast Fourier transform (FFT) and Lomb-Scargle periodograms. Figure 17 illustrates examples of these estimates. It should be noted that, despite being derived from the same recording, the estimates generated by the two methods show significant discrepancies. Chapter 6 demonstrates how these estimates change when faced with missing data along with robust correction methods.

Fourier-based spectral estimation

As Fourier-based methods require evenly sampled signals, estimations are performed on the heart rate modulating signal, $m(t)$, obtained from the Integral Pulse Frequency Modulation (IPFM) model [33]. This model assumes that the effects of the ANS on the SA node can be described by a

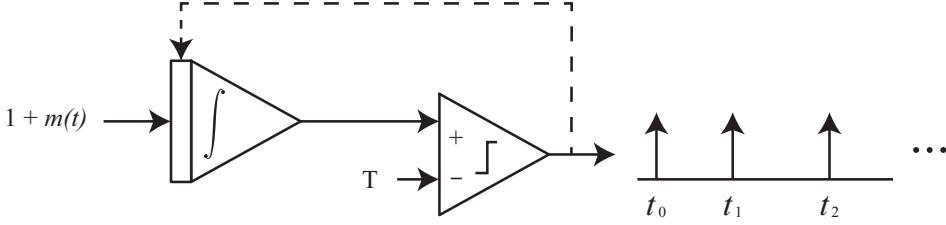


Figure 18: IPFM model diagram. The input function $m(t)$ modulates the variability, resulting in the event series t_k .

band-limited, zero-mean signal $m(t)$, with pulses triggered whenever the integral of this function reaches a predefined threshold (see Figure 18). Given a heartbeat time occurrence series (t_k) , the IPFM model relates the series with $m(t)$ as:

$$k = \int_0^{t_k} \frac{1 + m(\tau)}{T} d\tau, \quad (4.18)$$

k being the number of the current heartbeat and T the mean heart rate. According to [34], this technique produces more accurate spectra compared to those obtained from the interpolated R-R series using traditional interpolation methods such as cubic splines, thus avoiding spurious components and low-pass filtering effects. Furthermore, it allows the interpolation of the R-R series even under missing data conditions, as detailed in [31]. The major drawback of this method is the use of a constant heart rate. Although during short time periods the assumption of stationary conditions can be made, it is not realistic during long periods. In addition, stationarity can hardly be assumed in a context of wearable monitoring, where the user can perform a wide range of activities during a day. To overcome this, a version of the method known as Time-Varying IPFM (TVIPFM) model has been used [35, 36]. This modifies Equation 4.18 into

$$k = \int_0^{t_k} \frac{1 + m(\tau)}{T(\tau)} d\tau = \int_0^{t_k} d_{\text{HR}}(\tau) d\tau, \quad (4.19)$$

where $d_{\text{HR}}(t)$ is the instantaneous heart rate. Note that $T \rightarrow T(t)$. In this approach, k is generalized to continuous time, being $k(t)$ the continuous beat-

order function such that $k(t_k) = k$, and $d_{\text{HR}}(t)$ is obtained by differentiating $k(t)$ with respect to t [35]. Then, $m(t)$ is computed as

$$m(t) = \frac{d_{\text{HR}}(t) - d_{\text{mHR}}(t)}{d_{\text{mHR}}(t)}, \quad (4.20)$$

$d_{\text{mHR}}(t)$ representing the slow variations of the heart rate. This signal is obtained by low-pass filtering $d_{\text{HR}}(t)$ at 0.04 Hz.

Assuming stationarity in the analyzed segments, $m(t)$ spectral density is estimated using Welch's method, a non-parametric method that estimates a signal PSD by means of modified periodogram averaging. This technique reduces the variance of the estimate while sacrificing spectral resolution. First, $m(t)$ is sampled at 4 Hz, obtaining $m[n]$. For a segment $m_i[n]$ of N samples extracted from the signal $m[n]$, the periodogram $S_i(f)$ is calculated as the Fourier transform of the biased estimate of the autocorrelation sequence:

$$S_i(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} m_i[n] e^{-j2\pi f n} \right|^2. \quad (4.21)$$

A modified periodogram is an adapted version of the periodogram that aims to reduce spectral leakage by minimizing discontinuities at the boundaries of the data segment. This is achieved by applying a window function $w(n)$ of length N samples, same as the segment, to the signal segment before computing the Fourier transform. Thus, the modified periodogram $S_i^{\text{mod}}(f)$ is calculated as

$$S_i^{\text{mod}}(f) = \frac{1}{NW} \left| \sum_{n=0}^{N-1} w[n] m_i[n] e^{-j2\pi f n} \right|^2, \quad (4.22)$$

where W is a normalizing factor equal to the energy of $w(n)$. The average is calculated by generating N_{per} modified periodograms for each signal using sliding Hamming windows with a certain overlap:

$$S_{\text{Welch}}(f) = \frac{1}{N_{\text{per}}} \sum_{i=1}^{N_{\text{per}}} S_i^{\text{mod}}(f). \quad (4.23)$$

Averaged Lomb-Scargle periodograms

Averaged Lomb-Scargle PSD estimation makes use of Lomb-Scargle periodograms, a nonparametric spectral estimation based on a least-squares fit of sinusoids in data samples [37]. The main advantage of this method is that it can be used directly on unevenly sampled interval series, even in the presence of missing data. Estimates on heart rate representations are shown to be more accurate than in beat interval representations [34]. Therefore, Lomb-Scargle periodograms are computed on the inverse interval function $d_{\text{IIF}}(t_k)$, defined as:

$$d_{\text{IIF}}(t) = \sum_{k=1}^K \frac{1}{(t_k - t_{k-1})} \delta(t - t_k). \quad (4.24)$$

Thus, Lomb-Scargle periodograms are computed as:

$$S_{\text{Lomb}}(f) = \frac{[\sum_{k=1}^K (d_{\text{IIF}}(t_k) - \overline{d_{\text{IIF}}}) \cos(2\pi f(t_k - \tau))]^2}{2\sigma^2 \sum_{k=1}^K \cos^2(2\pi f(t_k - \tau))} + \frac{[\sum_{k=1}^K (d_{\text{IIF}}(t_k) - \overline{d_{\text{IIF}}}) \sin(2\pi f(t_k - \tau))]^2}{2\sigma^2 \sum_{k=1}^K \sin^2(2\pi f(t_k - \tau))}, \quad (4.25)$$

where the value of τ is defined as

$$\tan(4\pi f\tau) = \frac{\sum_{k=1}^K \sin(4\pi f t_k)}{\sum_{k=1}^K \cos(4\pi f t_k)}, \quad (4.26)$$

being $\overline{d_{\text{IIF}}}$ and σ^2 the mean and variance estimates of $d_{\text{IIF}}(t_k)$ respectively.

6.3 Frequency domain (II). Component separation

The previously established methods facilitate the analysis of different frequency components of variability. As discussed in Chapter 2, three primary frequency bands can be distinguished: VLF, LF, and HF. Traditionally, fixed frequency intervals have been employed: 0 to 0.04 Hz for VLF, 0.04 to 0.15 Hz for LF, and 0.15 to 0.4 Hz for HF. However, the validity of these intervals has been called into question [38–40]. The RSA component is not consistently confined to the classical HF band. It frequently overlaps with the LF component and may exhibit power beyond 0.4 Hz.

In response to this limitation, the scientific community has developed methods for more precise spectral separation. These approaches often incorporate respiratory information to more accurately localize the RSA component within the spectrum or to fully disentangle it from nonrespiratory components. The drawback of these methods, however, is their reliance on the availability of respiratory data, which is not always accessible. While it is sometimes possible to infer this information from ECG or PPG signals, the robustness of such inferences is highly dependent on signal quality.

For this reason, the traditional frequency windows remain widely used as they are effective in many applications. In this thesis, both classical frequency intervals and respiratory-driven component separation approaches are used.

Band selection using respiratory rate estimated from variability

One of the most straightforward methods for accurately assessing the RSA component involves estimating the respiratory frequency from HRV or PRV. This estimation can be performed using the spectral density of the respiratory signal, if such data are available, or inferred from the variability signal when the respiratory signal is unavailable. In a wearable environment, where access to the respiratory signal is uncommon, the latter approach is typically required.

A critical consideration with this method is the assumption of a distinct and stationary respiratory component, which holds in relaxed conditions,

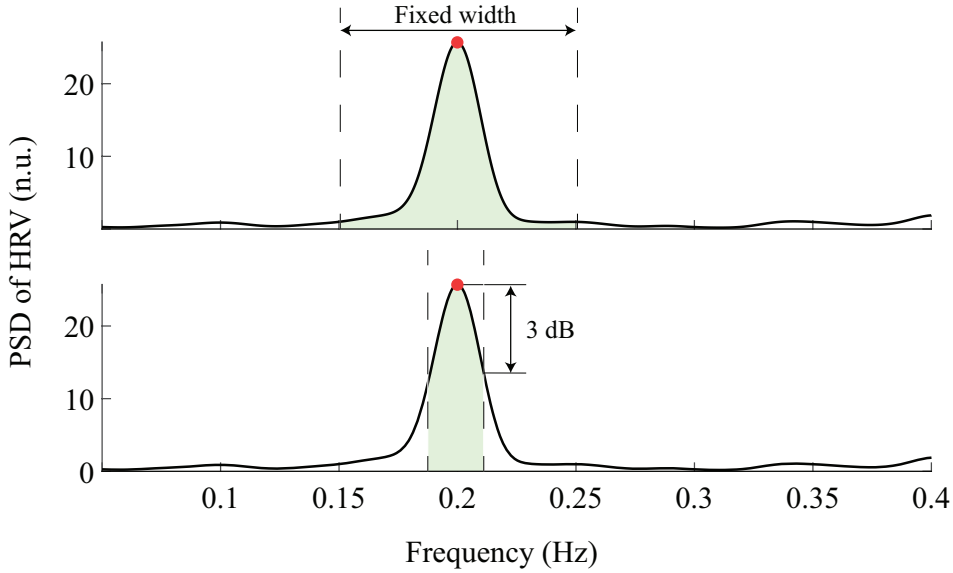


Figure 19: RSA band delimitation based in respiratory rate. Bands are centered in the respiratory rate (red dot), estimated from respiratory signals or other signals with respiratory modulation. *Top*: Fixed width. In this example, the bandwidth is 0.1 Hz. *Bottom*: -3 dB bandwidth.

such as when the user is silent or during controlled-breathing experiments. In such cases, respiratory frequency estimation is achieved by identifying the peak in the PSD of variability within a predefined range of expected frequencies.

However, in practical applications, the respiratory component often exhibits significant power at several frequencies, particularly when the user is speaking or engaged in different activities, making this method less suitable for continuous monitoring in wearable devices. To enhance the robustness of this approach, an initial evaluation of the sharpness of the spectrum can be incorporated, with estimation being performed only if the spectrum exceeds a defined sharpness threshold.

In this thesis, the approach without sharpness detection is employed, as is applied in a controlled breathing experiment (see Chapter 5). Under these conditions, a single-component spectrum is expected. The metronome respiratory frequencies used in this experiment are lower than those in

the classic HF band, *i.e.* below 0.15 Hz, making it necessary to center the respiratory band in the appropriate frequencies. Once the center frequency of the band is determined, the lower and upper limits are established using two methods: fixed bandwidths and -3 dB bandwidth (see Figure 19).

Spectral coherence

RSA band identification using spectral coherence depends on the variability signal combined with a respiratory signal. This signal may be true respiration or a proxy created from ECG or PPG signals, significantly influenced by breathing patterns, such as slope transit time or the width of the main wave [41, 42]. Spectral coherence, $C_{xy}(f)$, measures the linear correlation between the two signals $x(t)$ and $y(t)$ at various frequencies. It takes real values between 0 and 1, with higher values indicating a stronger correlation at a given frequency. Formally, it is defined as

$$C_{xy}(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)}, \quad (4.27)$$

where $S_{xx}(f)$ and $S_{yy}(f)$ represent the spectral densities of the signals $x(t)$ and $y(t)$, respectively, estimated using periodograms (see Equation 4.21). $S_{xy}(f)$ denotes the cross-spectral density, which is the Fourier transform of the cross-correlation between the two signals.

Contiguous frequencies exceeding a certain threshold around the estimated respiratory rate are selected to delineate the RSA band (see Figure 20). This threshold is determined as the 95th percentile of one hundred realizations of Gaussian white noise, ensuring that the observed coherence is statistically significant. Additional conditions may be imposed to ensure robust evaluation. This thesis excluded frequency bands narrower than 0.015 Hz.

Orthogonal Subspace Projection

The last reviewed method is Orthogonal Subspace Projection (OSP) [36, 43]. In this approach $m[n]$ is decomposed by the respiratory signal, $r[n]$, sampled at the same points, into a component linearly related to respiration, $\hat{m}_r[n]$,

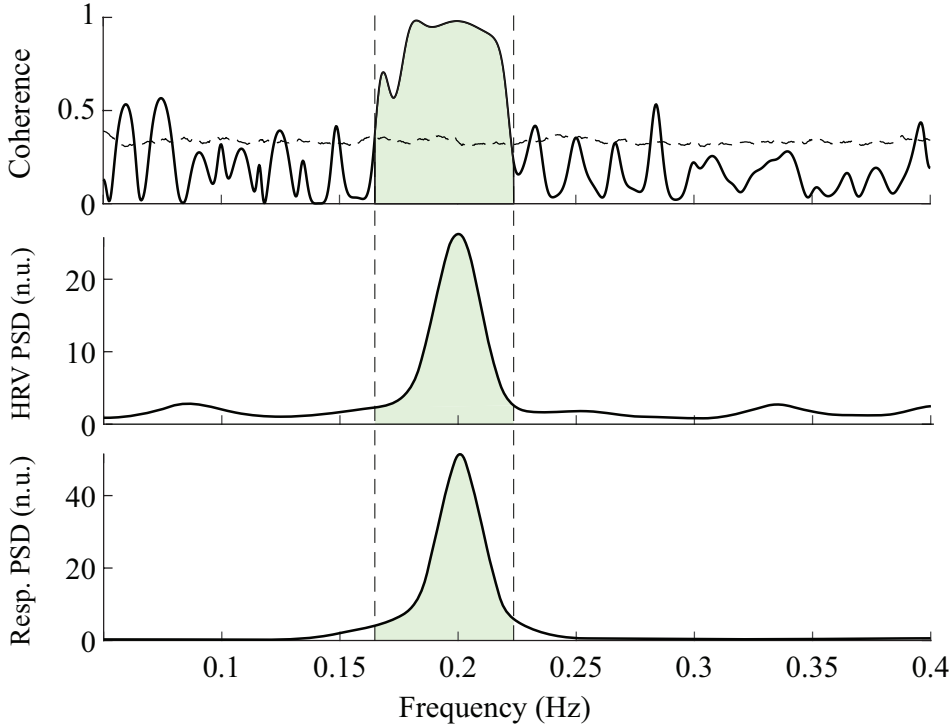


Figure 20: RSA band delimitation based in spectral coherence. *Top:* Spectral coherence between HRV and a respiration-modulated signal. The dashed line represents the threshold that defines the RSA band. *Center:* HRV spectral density. *Bottom:* Spectral density of a respiration-modulated signal. In this example, the slope transit time (see Chapter 5).

and a residual component, $\hat{m}_\perp[n]$, which contains other modulators. Unlike previous methods, this approach does not focus on dividing a single PSD into frequency bands; rather, it separates the modulating signal into two distinct signals to estimate two separate PSDs.

OSP is used to decompose a signal when provided with a reference, in this case a respiratory reference. The initial step involves the construction of a subspace characterized by the basis \mathbf{V} , which is indicative of respiratory activity. OSP was originally suggested in [44] for the decomposition of the tachogram and used wavelet transforms applied to the respiration signal to establish \mathbf{V} . In this thesis, delayed versions of $r[n]$ were used instead.

Considering $m[n]$ and $r[n]$ signals consisting of N samples and p as the order

of the model, $m[n]$ can be projected onto a subspace V with dimensions $(N - p + 1) \times p$, defined by all variations in $r[n]$:

$$V = \begin{bmatrix} r[0] & r[1] & \dots & r[p-1] \\ r[1] & r[2] & \dots & r[p] \\ \vdots & \vdots & \ddots & \vdots \\ r[N-p] & r[N-p+1] & \dots & r[N-1] \end{bmatrix}. \quad (4.28)$$

A projection matrix P with dimensions $(N - p + 1) \times (N - p + 1)$ is created as

$$P = V(V^T V)^{-1} V^T. \quad (4.29)$$

Then $m = \{m[p-2], m[p+1] \dots m[N-1]\}$ is composed, and \hat{m}_r and \hat{m}_\perp are computed as

$$\hat{m}_r^T = P m^T \quad (4.30)$$

and

$$\hat{m}_\perp = m - \hat{m}_r. \quad (4.31)$$

Note that $\hat{m}_r[n]$ and $\hat{m}_\perp[n]$ consist of $N - p + 1$ samples, since $p + 1$ samples are required to compose V . This approach is analogous to fitting an autoregressive moving average model that describes the interactions between $m[n]$ and $r[n]$. Refer to [36] for a comprehensive discussion on this topic. Once these signals are obtained, their PSDs, $\hat{S}_{m_r}(f)$ and $\hat{S}_{m_\perp}(f)$, can be estimated by any of the methods described above. Figure 21 illustrates the OSP decomposition of $m[n]$, presenting one scenario where the respiratory rate exceeds the LF range and another where it is within the LF range. Observe that the former scenario is expected to yield results comparable to the conventional fixed LF and HF ranges, whereas the latter scenario results in significantly different outcomes.

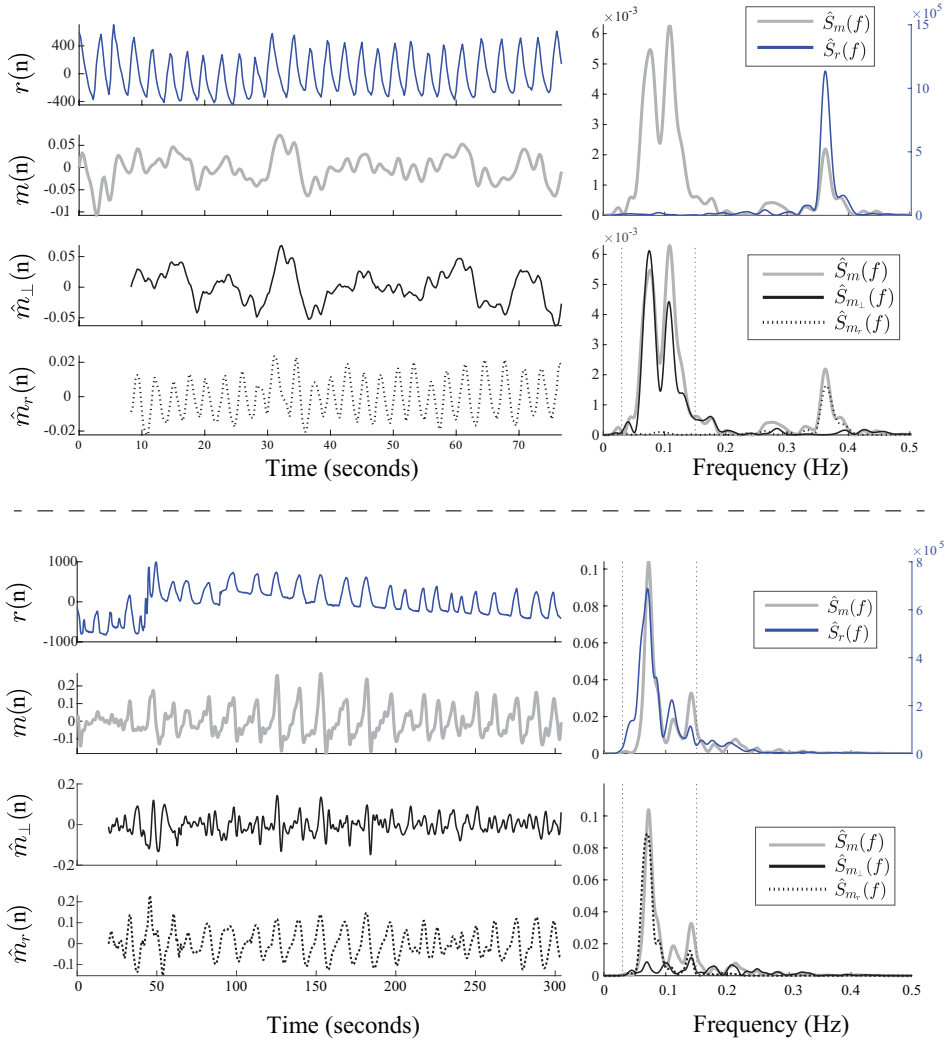


Figure 21: OSP decomposition. *Top panel:* Respiratory rate outside LF band. *Bottom panel:* Respiratory rate within LF band. *Inside each panel, left side top to bottom:* Respiratory signal, modulation of the heart rate, modulation linearly related with respiration, and residual component. *Inside each panel, right side top:* Respiratory signal (blue) and modulation of the heart rate (gray) spectra. *Inside each panel, right side bottom:* Modulation of the heart rate (gray), modulation linearly related with respiration (black), and residual component (dotted) spectra.

6.4 Frequency domain (III). Metrics computation

The main metrics in frequency-domain HRV analysis include:

- The low-frequency power (P_{LF}), that represents both sympathetic and parasympathetic activity of the ANS.
- The high-frequency power (P_{HF}), primarily associated with parasympathetic activity and closely related to respiration.
- The LF power measured in normalized units (P_{LFn}), used to emphasize the balance between the LF and HF components of HRV, providing a clearer picture of the relative contributions of sympathetic and parasympathetic activity, independent of the overall variability.
- The P_{LF}/P_{HF} ratio. This ratio has been widely used as an indicator of sympathetic-parasympathetic balance. A high value suggests sympathetic dominance, while a low value indicates parasympathetic dominance. However, this metric has been subject to criticism as an indicator of sympathovagal balance [45, 46], and exhibits a lack of robustness when P_{HF} approaches values close to zero.

P_{LF} and P_{HF} are determined by applying trapezoidal integration to the PSD within their respective frequency ranges. P_{LF} is consistently integrated over the interval 0.04 to 0.15 Hz, whereas the HF band's boundaries can either be the classic window of 0.15 to 0.4 Hz; or adjusted based on respiration guidance, as explained before. P_{LFn} is computed as $P_{LF}/(P_{LF}+P_{HF})$.

OSP decomposition method requires adapted frequency-domain metrics. Spectral powers $P_{r,LF+HF}$, representing the parasympathetic nervous system, and $P_{\perp,LF}$, representing the sympathetic, are calculated by integrating $\hat{S}_{m_r}(f)$ and $\hat{S}_{m_{\perp}}(f)$ over 0.04- F_{\max} Hz and 0.04-0.15 Hz, respectively, using trapezoidal numerical integration. F_{\max} is the upper limit of the spectrum, calculated as half of the MHR, which can be seen as the Nyquist criterion, since heart rate is inherently sampled beat to beat.

To address concerns about the traditional $P_{\text{LF}}/P_{\text{HF}}$ ratio as a measure of sympathovagal balance, a new decomposition-based version of this index was proposed in [43], described as

$$\mathcal{R} = \frac{P_{\perp, \text{LF}}}{P_{r, \text{LF} + \text{HF}}}, \quad (4.32)$$

since $P_{r, \text{LF} + \text{HF}}$ and $P_{\perp, \text{LF}}$ may be a more precise representation of parasympathetic and sympathetic activity than P_{HF} and P_{LF} [43, 45, 47]. However, when the coupling between cardiorespiratory signals is weak, \mathcal{R} may assume excessively large values [47]. This issue can be alleviated by normalizing \mathcal{R} to become

$$\mathcal{R}' = \frac{P_{\perp, \text{LF}}}{P_{r, \text{LF} + \text{HF}} + P_{\perp, \text{LF}}}. \quad (4.33)$$

\mathcal{R}' is analogous to the conventional normalized ratio, P_{LFn} , in which the spectral power within the LF band is normalized against the total spectral power across all combined bands [36].

6.5 Lagged Poincaré plots

Lagged Poincaré plots serve as graphical tools to illustrate interbeat dynamics. These graphs display variability by charting the IBI series against itself with a certain lag, usually 1, creating a scatterplot that outlines the phase space of the series (see Figure 22). A spread distribution of the points signifies increased variability. Poincaré plots have become essential in HRV analysis, proving to offer valuable prognostic insights into myocardial infarction, chronic heart failure, and sudden infant death syndrome [48]. Recent findings have validated the reliability of Poincaré plots even in ultra-short-term monitoring [49]. The main challenge with Poincaré plots lies in the absence of clear quantitative metrics to describe the key characteristics of the plot. In this context, the ellipse fitting method has been selected [48]. An ellipse is tailored to align with the point distribution, and geometric indices are derived. These indexes are as follows:

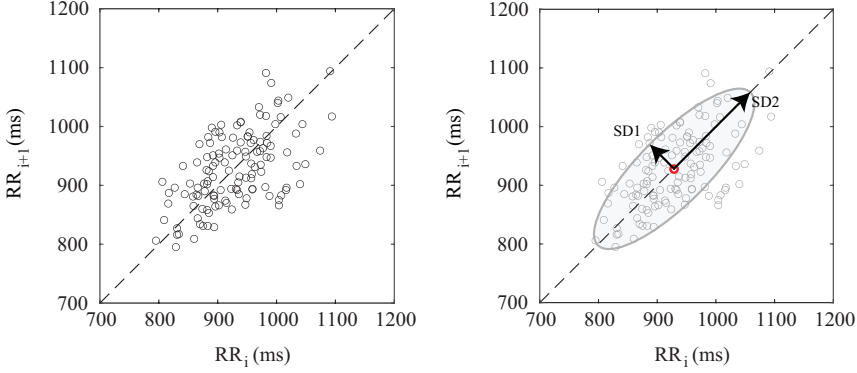


Figure 22: Lagged Poincaré plots. *Left:* Lagged Poincaré Plot with one-beat lag. *Right:* Ellipse fitting method.

- SD1: the standard deviation of the projection of points on the ellipse axis perpendicular to the line of identity. This measure is related to the fast beat-to-beat variability and RMSSD:

$$SD1 = \frac{RMSSD}{\sqrt{2}}. \quad (4.34)$$

- SD2: the standard deviation of the projected points along the line-of identity, that measures long term dynamics of HRV.
- SD1/SD2: the ratio, to describe the relationship between these components.
- $S = \pi \cdot SD1 \cdot SD2$: the area of the ellipse, related to overall variability. It is proportional with the variance of the N-N interval, and thus with $SDNN^2$
- Md: the mean data distance to the ellipse centroid (euclidean norm).
- Sd: the standard deviation to the ellipse centroid (euclidean norm).

SD1 and SD2 are also related to SDNN as

$$SDNN = \sqrt{\frac{SD1^2 + SD2^2}{2}}. \quad (4.35)$$

Bibliography

- [1] Natasa Reljin, Jesus Lazaro, Md Billal Hossain, Yeon Sik Noh, Chae Ho Cho, and Ki H. Chon. Using the Redundant Convolutional Encoder–Decoder to Denoise QRS Complexes in ECG Signals Recorded with an Armband Wearable Device. *Sensors*, 20(16):4611, 8 2020.
- [2] F.L. Mvuh, C.O.V. Ebode Ko’a, and B. Bodo. Multichannel high noise level ECG denoising based on adversarial deep learning. *Scientific Reports*, 14(1), 2024.
- [3] C. Manickam, M. Govindasamy, S. Muthusamy, and M. Paramasivam. A Novel Method for Design and Implementation of Systolic Associative Cascaded Variable Leaky Least Mean Square Adaptive Filter for Denoising of ECG Signals. *Wireless Personal Communications*, 137(2):1029–1043, 2024.
- [4] H. Shin. Signal completion using generative adversarial networks for enhanced photoplethysmography measurement accuracy. *Computers in Biology and Medicine*, 180, 2024.
- [5] R. Bailón, L. Sörnmo, and P. Laguna. A robust method for ECG-based estimation of the respiratory frequency during stress testing. *IEEE Transactions on Biomedical Engineering*, 53(7):1273–1285, 2006.
- [6] Bo Hjorth. EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3):306–310, 1 1970.
- [7] Elena Peralta, Jesus Lazaro, Raquel Bailon, Vaidotas Marozas, and Eduardo Gil. Optimal fiducial points for pulse rate variability analysis

- from forehead and finger photoplethysmographic signals. *Physiological Measurement*, 40(2), 2 2019.
- [8] J. Song, D. Li, X. Ma, G. Teng, and J. Wei. PQR signal quality indexes: A method for real-time photoplethysmogram signal quality estimation based on noise interferences. *Biomedical Signal Processing and Control*, 47:88–95, 2019.
- [9] Mohamed Elgendi. Optimal signal quality index for photoplethysmogram signals. *Bioengineering*, 3(21), 12 2016.
- [10] Christina Orphanidou, Timothy Bonnici, Peter Charlton, David Clifton, David Vallance, and Lionel Tarassenko. Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring. *IEEE Journal of Biomedical and Health Informatics*, 19(3):832–838, 5 2015.
- [11] Yogesh Gautam and Houtan Jebelli. Autoencoder-based Photoplethysmography (PPG) signal reliability enhancement in construction health monitoring. *Automation in Construction*, 165:105537, 9 2024.
- [12] Q. Li and G.D. Clifford. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiological Measurement*, 33(9):1491–1501, 2012.
- [13] Z. Wang, C. Ma, S. Zhang, Y. Li, L. Zhao, J. Li, and C. Liu. Improving ECG signal quality assessment through swin-transformer-based generative adversarial network. *Biomedical Signal Processing and Control*, 98, 2024.
- [14] Win-Ken Beh, Yu-Chia Yang, and An-Yeu Wu. Quality-Aware Signal Processing Mechanism of PPG Signal for Long-Term Heart Rate Monitoring. *Sensors*, 24(12):3901, 6 2024.
- [15] Juan Pablo Martínez, Rute Almeida, Salvador Olmos, Ana Paula Rocha, and Pablo Laguna. A Wavelet-Based ECG Delineator Evaluation on Standard Databases. *IEEE Transactions on Biomedical Engineering*, 51(4):570–581, 4 2004.

- [16] M. Liu, S. Zhao, Z. Zhang, J. Zhang, H. Du, X. Cao, and P. Xiong. *ECG Signal Delineation Based on Multi-scale Channel Attention Convolutional Neural Network*, volume 1163 LNEE. Springer, 2024.
- [17] Guillermo Jimenez-Perez, Juan Acosta, Alejandro Alcaine, and Oscar Camara. Generalising electrocardiogram detection and delineation: training convolutional neural networks with synthetic data augmentation. *Frontiers in Cardiovascular Medicine*, 11, 2024.
- [18] Philipp Sodmann, Marcus Vollmer, Neetika Nath, and Lars Kaderali. A convolutional neural network for ECG annotation as the basis for classification of cardiac rhythms. *Physiological Measurement*, 39(10):104005, 10 2018.
- [19] Geoffrey H. Tison, Jeffrey Zhang, Francesca N. Delling, and Rahul C. Deo. Automated and Interpretable Patient ECG Profiles for Disease Detection, Tracking, and Discovery. *Circulation: Cardiovascular Quality and Outcomes*, 12(9), 9 2019.
- [20] A. Peimankar and S. Puthusserypady. DENS-ECG: A deep learning approach for ECG signal delineation. *Expert Systems with Applications*, 165, 2021.
- [21] Siti Nurmaini, Annisa Darmawahyuni, Muhammad Naufal Rachmatullah, Jannes Effendi, Ade Iriani Sapitri, Firdaus Firdaus, and Bambang Tutuko. Beat-to-Beat Electrocardiogram Waveform Classification Based on a Stacked Convolutional and Bidirectional Long Short-Term Memory. *IEEE Access*, 9:92600–92613, 2021.
- [22] Xinyue Li, Wenjie Cai, Bolin Xu, Yupeng Jiang, Mengdi Qi, and Mingjie Wang. SEResUTer: a deep learning approach for accurate ECG signal delineation and atrial fibrillation detection. *Physiological Measurement*, 44(12):125005, 12 2023.
- [23] Duoduo Wang, Lishen Qiu, Wenliang Zhu, Yanfang Dong, Huimin Zhang, Yuhang Chen, and lirong Wang. Inter-patient ECG characteristic wave detection based on convolutional neural network combined

- with transformer. *Biomedical Signal Processing and Control*, 81:104436, 3 2023.
- [24] Guillermo Jimenez-Perez, Alejandro Alcaine, and Oscar Camara. Delineation of the electrocardiogram with a mixed-quality-annotations dataset using convolutional neural networks. *Scientific Reports*, 11(1):863, 1 2021.
- [25] Jesús Lázaro, Eduardo Gil, José María Vergara, and Pablo Laguna. Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children. *IEEE Journal of Biomedical and Health Informatics*, 18(1):240–246, 2014.
- [26] P.H. Charlton, K. Kotzen, E. Mejía-Mejía, P.J. Aston, K. Budidha, J. Mant, C. Pettit, J.A. Behar, and P.A. Kyriacou. Detecting beats in the photoplethysmogram: Benchmarking open-source algorithms. *Physiological Measurement*, 43(8), 2022.
- [27] K. Kotzen, P.H. Charlton, A. Landesberg, and J.A. Behar. Benchmarking Photoplethysmography Peak Detection Algorithms Using the Electrocardiogram Signal as a Reference. In *Computing in Cardiology*, volume 2021-Septe, 2021.
- [28] H. Luczak and W. Laurig. An Analysis of Heart Rate Variability. *Ergonomics*, 16(1):85–97, 1973.
- [29] O Rompelman, A J R M Coenen, and R I Kitney. Measurement of heart-rate variability: Part I-Comparative study of heart.rate variability analysis methods. *Biol. Eng. & Comput*, 15:233–239, 1977.
- [30] Hussein Al Osman, Mohamad Eid, and Abdulmotalieb El Saddik. A Pattern-Based Windowed Impulse Rejection Filter for Nonpathological HRV Artifacts Correction. *IEEE Transactions on Instrumentation and Measurement*, 64(7):1944–1957, 7 2015.

- [31] Javier Mateo and Pablo Laguna. Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal. *IEEE Transactions on Biomedical Engineering*, 50(3):334–343, 3 2003.
- [32] James McNames and Mateo Aboy. Reliability and accuracy of heart rate variability metrics versus ECG segment duration. *Medical & Biological Engineering & Computing*, 44(9):747–756, 9 2006.
- [33] Elliott J Bayly. Spectral Analysis of Pulse Frequency Modulation in the Nervous Systems. *IEEE Transactions on Biomedical Engineering*, 15(4):257–265, 1968.
- [34] Javier Mateo and Pablo Laguna. Improved Heart Rate Variability Signal Analysis from the Beat Occurrence Times According to the IPFM Model. *IEEE Transactions on Biomedical Engineering*, 47(8), 2000.
- [35] R. Bailón, G. Laouini, C. Grao, M. Orini, P. Laguna, and O. Meste. The integral pulse frequency modulation model with time-varying threshold: Application to heart rate variability analysis during exercise stress testing. *IEEE Transactions on Biomedical Engineering*, 58(3 PART 1):642–652, 2011.
- [36] Leif Sornmo, Raquel Bailon, and Pablo Laguna. Spectral Analysis of Heart Rate Variability in Time-Varying Conditions and in the Presence of Confounding Factors. *IEEE Reviews in Biomedical Engineering*, 2024.
- [37] Pablo Laguna, George B Moody, Associate Member, Roger G Mark, and Senior Member. Power Spectral Density of Unevenly Sampled Data by Least-Square Analysis: Performance and Application to Heart Rate Signals. *IEEE Transactions on Biomedical Engineering*, 45(6):698–715, 1998.
- [38] B. Aysin and E. Aysin. Effect of respiration in heart rate variability (HRV) analysis. In *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, pages 1776–1779, 2006.
- [39] Alberto Hernando, Jesús Lázaro, Eduardo Gil, Adriana Arza, Jorge Mario Garzón, Raúl López-Antón, Concepción De La Camara,

- Pablo Laguna, Jordi Aguiló, and Raquel Bailón. Inclusion of Respiratory Frequency Information in Heart Rate Variability Analysis for Stress Assessment. *IEEE Journal of Biomedical and Health Informatics*, 20(4):1016–1025, 7 2016.
- [40] M. Daoud, P. Ravier, and O. Buttelli. Use of cardiorespiratory coherence to separate spectral bands of the heart rate variability. *Biomedical Signal Processing and Control*, 46:260–267, 2018.
- [41] M.D. Peláez-Coca, M. Orini, J. Lázaro, R. Bailón, and E. Gil. Cross time-frequency analysis for combining information of several sources: Application to estimation of spontaneous respiratory rate from photoplethysmography. *Computational and Mathematical Methods in Medicine*, 2013, 2013.
- [42] Jesús Lázaro, Yunyoung Nam, Eduardo Gil, Pablo Laguna, and Ki H. Chon. Respiratory rate derived from smartphone-camera-acquired pulse photoplethysmographic signals. *Physiological Measurement*, 36(11):2317–2333, 10 2015.
- [43] Carolina Varon, Jesus Lazaro, Juan Bolea, Alberto Hernando, Jordi Aguiló, Eduardo Gil, Sabine Van Huffel, and Raquel Bailón. Unconstrained Estimation of HRV Indices after Removing Respiratory Influences from Heart Rate. *IEEE Journal of Biomedical and Health Informatics*, 23(6):2386–2397, 11 2019.
- [44] Devy Widjaja, Alexander Caicedo, Elke Vlemincx, Ilse Van Diest, and Sabine Van Huffel. Separation of respiratory influences from the tachogram: A methodological evaluation. *PLoS ONE*, 9(7), 7 2014.
- [45] D.L. Eckberg. Sympathovagal balance: A critical appraisal. *Circulation*, 96(9):3224–3232, 1997.
- [46] George E. Billman. The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. *Frontiers in Physiology*, 4, 2013.
- [47] Binbin Liu, Saisai Yan, Xiaoni Wang, Lin Xie, Jie Tong, Fadong Zhao, Xiaohui Di, Xiangguo Yan, and Jianbao Zhang. An improved method to

- evaluate heart rate variability based on time-variant cardiorespiratory relation. *Journal of Applied Physiology*, 127(2):320–327, 8 2019.
- [48] Berik Koichubekov, Viktor Riklefs, Marina Sorokina, Ilya Korshukov, Lyudmila Turgunova, Yelena Laryushina, Rishan Bakirova, Gulmira Muldaeva, Ernur Bekov, and Makhabbat Kultenova. Informative nature and nonlinearity of lagged poincaré plots indices in analysis of heart rate variability. *Entropy*, 19(10), 10 2017.
- [49] Mimma Nardelli, Alberto Greco, Juan Bolea, Gaetano Valenza, Enzo Pasquale Scilingo, and Raquel Bailon. Reliability of lagged poincaré plot parameters in ultrashort heart rate variability series: Application on affective sounds. *IEEE Journal of Biomedical and Health Informatics*, 22(3):741–749, 5 2018.

Respiration-guided ANS evaluation from wrist PVP

1 Objectives and motivations

This study continues the work of Hernando *et al.*, who evaluated the utility of the non-invasively acquired PVP signal as an indicator of the ANS [1, 2], which was preceded by investigations that used the signal obtained invasively in pigs [3]. Hernando *et al.* demonstrated that PVP measured from NIVA, that actually measures changes in PVP (*i.e.*, a pressure derivative), in the volar aspect of the wrist and PPG in the same place (see Chapter 2), is a reliable surrogate for HRV under resting conditions. In addition, they found a significantly higher power within the HF band at PVP, defined as the band centered at the respiratory rate. It suggests that PVP may provide improved respiratory rate information, perhaps because venous return is strongly mediated by negative pressure in the chest caused by inspiratory movements.

HRV analysis can better reflect vagal modulation of heart rate by including respiratory information, as has been reported for some applications, including stress assessment [4]. Thus, NIVA could make use of its novel signal to obtain enhanced information on parasympathetic tone guided by respiration, which could be used in stress and well-being identification

applications (*e.g.*, yoga, mindfulness, etc.) using a wrist device, the most accepted on the market nowadays.

The purpose of this research is to further investigate the PVP signal obtained from NIVA, examining its ability to extract respiration-driven parasympathetic indicators. Respiratory data and pulse rate have been extracted from the signal and subsequently used to calculate the PRV power in a respiration-dependent frequency band. Performance has been evaluated in a breathing-controlled experiment with different respiratory rates. It is known that, for the same subject, the power in the band defined by respiration tends to decrease with an increase in the respiratory rate [5]. The hypothesis is that the PVP signal could be more sensitive than the PPG signal to these intra-subject variations.

2 Materials and methods

In collaboration with Vanderbilt University Medical Center and under COI IRB approval, eight healthy volunteers were recorded with the NIVA device, acquiring PVP and PPG signals while maintaining controlled respiration at 0.1, 0.2, and 0.3 Hz (6, 12, and 18 bpm, respectively). Each stage lasted 5 to 6 minutes, and the subjects were seated throughout the experiment. The PVP signal is acquired using a piezoelectric sensor that samples at 500 Hz, while the PPG uses an optoelectric one at 125 Hz, both in the volar aspect of the wrist. It is important to note that the PVP obtained by NIVA is a signal proportional to the pressure derivative, rather than to the pressure itself. For simplicity, this signal is referred to as PVP rather than PVP' or similar, as pressure information is contained in it. However, this fact is taken into account in the processing explained below.

2.1 Signal delineation

All signals were low-pass filtered with a cutoff frequency of 5 Hz for PVP and 2 Hz for PPG. These frequencies were chosen empirically to eliminate high-

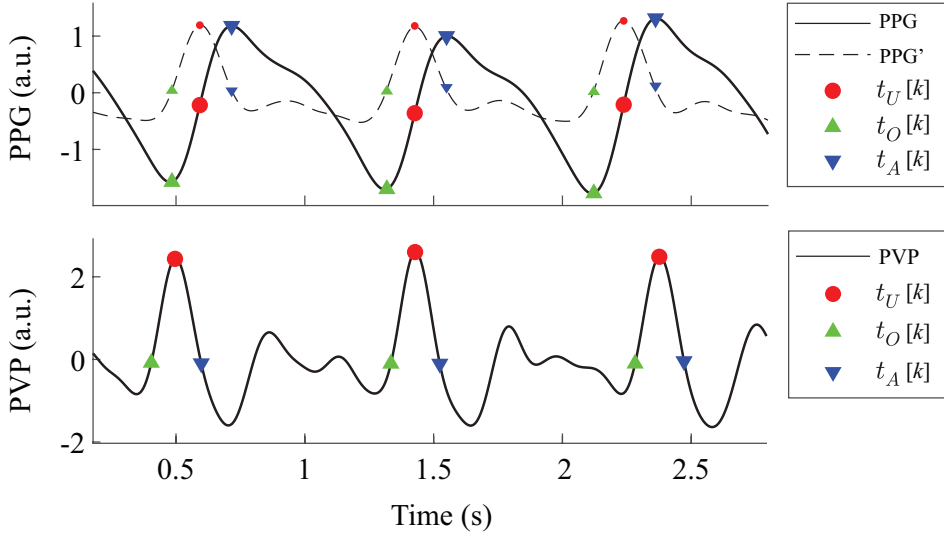


Figure 23: PPG and PVP delineation. Key points include the maximum upslope ($t_U[k]$) marked by red circles, the pulse onset ($t_O[k]$) marked by green triangles, and the derivative zero crossing ($t_A[k]$) marked by blue triangles. The PPG signal is represented by a solid line, while its derivative is shown with a dashed line. *Bottom:* PVP delineation.

frequency noise. The cutoff frequency for PPG is lower than typical values for wrist PPGs due to the poor signal-to-noise ratio caused by prioritizing the piezoelectric placement over the vein, which led to suboptimal positioning of the optical sensor. For a better comparison of the signals, the results have also been calculated for PVP filtered at 2 Hz. The signals were high-pass filtered with a cutoff frequency of 0.5 Hz to remove low-frequency oscillations such as baseline wander.

PPG and PVP are delineated to obtain key points (see Figure 23). In the case of PPG, delineation is applied to the derivative obtained with a differentiation filter. PVP obtained by NIVA is proportional to the pressure derivative; thus, the key points are delineated without this differentiation process. First, the maximum upslope of each pulse, $t_U[k]$, is obtained using an adaptive threshold, with $k = 1, 2, \dots, K$, where K represents the total number of pulses. Two more key points are obtained: the signal onset, denoted as $t_O[k]$; and the derivative zero crossing, denoted as $t_A[k]$. The calculation of $t_O[k]$ is carried out as described in [6]. In PVP, the derivative zero crossing

is the zero crossing of the signal itself.

2.2 Respiratory rate estimation

The respiratory frequency was estimated through spectral analysis of both Slope Transit Time (STT) and PRV. STT is defined as the time interval between the PPG waveform onset ($t_O[k]$) and its maximum point ($t_A[k]$). This definition has been extended to the PVP using the same key points.

STT is influenced by respiration through several physiological mechanisms, such as respiratory-induced variations in stroke volume, RSA, and changes in peripheral vascular tone. In this way, respiratory-induced variations in STT can be used to estimate respiratory rate, analogously to the PRV. The STT signal, $d_{STT}(t)$, is computed as

$$d_{STT}(t) = \sum_{k=1}^K (t_A[k] - t_O[k]) \delta(t - t_A[k]). \quad (5.1)$$

A running standard deviation threshold was used as an outlier detector (see Equation 4.5). This outlier detector had a minimum and maximum threshold to eliminate both false negatives and false positives in the delineation. Subsequently, the corrected $d_{STT}(t)$ was interpolated using cubic splines at 4 Hz to ensure uniform sampling, obtaining $d_{STT}[n]$.

PRV is computed from $t_U[k]$ using the IPFM model, thus obtaining a heart modulation signal $m[n]$ at a sampling rate of 4 Hz, consistent with $d_{STT}[n]$. Outliers in $m[n]$ are less common, as $m[n]$ computation depends on the maximum upslope, which is more robust than the other two key points. Outliers in the $t_U[k]$ series are treated as ectopic beats in the computation of $m[n]$ with the IPFM model, following the method described in [7].

The respiratory rate was determined by identifying the frequency with the highest power within the range of 0.05 to 0.4 Hz in the PSD of both $m[n]$ and $d_{STT}[n]$, obtaining separated estimates. PSD was calculated using the Welch method, employing a 60-second Hamming window with 50% overlap. In

the absence of a direct respiratory signal for validation, a criterion was established to assess the precision of the estimated respiratory frequency. An estimate was deemed accurate if it fell within ± 0.03 Hz of the metronome-guided breathing frequency. This tolerance accounts for potential minor deviations in the participant's adherence to the prescribed breathing pattern. The efficacy of the respiratory rate estimation method was quantified using a success rate metric, separately for $m[n]$ and $d_{\text{STT}}[n]$ estimations. This metric is defined as the ratio of the correct estimates (those that meet the ± 0.03 Hz criterion) to the total number of cases examined, expressed as a proportion.

2.3 Respiration-guided parasympathetic assessment

Three different methods have been used to calculate the RSA/HF bounds in the $m[n]$ spectra. The first one consists of 0.1 Hz width windows centered around the estimated respiratory rate. The second method centers the band in a similar manner but adjusts the bandwidth to -3 dB instead of using a fixed width. The third is based on the spectral coherence between the $m[n]$ and $d_{\text{STT}}[n]$ spectra. The boundaries of the HF window are examined by identifying frequencies with coherence that exceed a threshold near the estimated respiratory rate. This threshold was previously calculated to be the 95th percentile of a hundred instances of white Gaussian noise, ensuring that the coherence is statistically meaningful. Windows less than 0.015 Hz wide are rejected.

After determining the window boundaries, the HF power (P_{HF}) is computed from the $m[n]$ spectrum using trapezoidal integration. The spectra are normalized by dividing by the power integral in the range of 0.03 to 0.4 Hz. A paired signed rank test is performed, in which the hypothesis of zero median is tested for P_{HF} distributions at different respiratory rates. Thus, parasympathetic estimates are compared, obtaining a qualitative measurement of discrimination between different breathing rates of PPG and PVP. A significance level of $\alpha = 0.05$ is used to reject or accept the null hypothesis.

3 Results

Respiratory rate estimation results are shown in Table 2. A 100% success rate is achieved using the PVP's $m[n]$ spectrum when filtered at both 5 Hz and 2 Hz. The corresponding root mean square errors (RMSE) compared to the metronome rate are 0.7 mHz at 5 Hz and 1.1 mHz at 2 Hz. When using the PPG, the success rate decreases to 87.5%, although the RMSE remains low at 1.3 mHz in the successful cases.

In contrast, the outcomes with the $d_{\text{STT}}[n]$ spectrum are less favorable. For the filtered PVP, a success rate of 66.7% is obtained at both 5 Hz and 2 Hz, with RMSE values of 6 mHz and 6.8 mHz, respectively. In the case of PPG, the success rate is 75% with an RMSE of 1 mHz.

Regarding respiration-guided parasympathetic assessment, downward trends in P_{HF} are present in both PVP and PPG when using fixed-width windows centered at each respiratory rate. There are significant differences ($p < 0.05$) in the median between the 6-12 bpm and 12-18 bpm distributions for both PPG and PVP (see Figure 24). Similar results are obtained using -3 dB wide windows for PVP. However, the differences are no longer significant for PPG. The utilization of coherence-defined windows proves ineffective for both PVP and PPG.

Table 2: Respiratory rate estimation from PVP and PPG.

Spectrum	Sensor	Filter cutoff (Hz)	Success rate (%)	RMSE (mHz)
$m[n]$	PVP	5	100.0	0.7
		2	100.0	1.1
	PPG	2	87.5	1.3
$d_{\text{STT}}[n]$	PVP	5	66.7	6.0
		2	66.7	6.8
	PPG	2	75.0	1.0

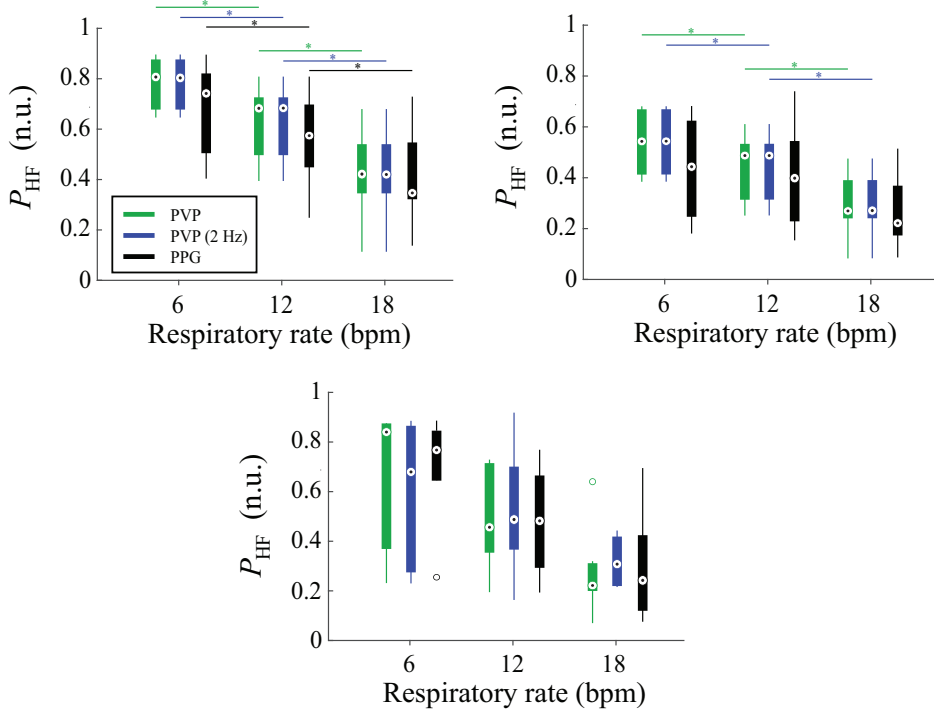


Figure 24: Signed rank test results. The test measures differences in median distributions of P_{HF} at different respiratory rates (6, 12 and 18 breaths per minute). Statistical significance is marked with an asterisk (*). *Top left:* Fixed width HF window, centered on respiratory rate. *Top right:* -3 dB width HF window, centered on respiratory rate. *Bottom:* Coherence-based HF window.

4 Discussion

4.1 Findings

The PVP signal is strongly influenced by breathing, making it especially useful for estimating respiratory rate. In fact, using $m[n]$ derived from the PVP signal achieved 100% accuracy in estimating respiratory rate, outperforming $m[n]$ derived from wrist PPG.

Meanwhile, $d_{STT}[n]$ performed better with PPG than with PVP, although it did not surpass the accuracy achieved by $m[n]$. This information was

used to accurately delimit the HF band. In this task, both PVP and PPG were able to distinguish between groups at different respiratory rates using fixed-width windows. However, only PVP worked effectively with -3 dB width windows. Neither PVP nor PPG was able to distinguish between groups using coherence-based windows. This may be because $d_{\text{STT}}[n]$ was not as good an estimator of respiration as $m[n]$ in this particular application.

In this research, the two signals offered by the device were compared; however, the combination of both could give additional information that could not be achieved with the signals separately. This is a future line to consider together with comparing performances also in a spontaneous breathing scenario.

4.2 Limitations

The respiratory rate has been obtained from the $m[n]$ spectrum instead of from the raw signal spectrum, as in [1], because in the PVP signal the respiration band was heavily filtered, probably due to the high-pass nature of the derivative. In the absence of a respiratory signal, the metronome frequency was used as a reference, since in [8] it was shown that the deviations are of the order of tenths of millihertz.

Another issue to be addressed is that the tidal volume was not measured during the experiment. Changes in respiratory rate are expected to cause changes in tidal volume, and variations in tidal volume are known to affect P_{HF} in the opposite way as changing respiratory rate. The higher the tidal volume, the greater P_{HF} and vice versa [5]. This variation should be considered in future experiments.

Finally, the size of the database is a clear limitation. A larger number of subjects should be involved in future research.

5 Conclusion

In this work, the ability to derive respiration-guided parasympathetic estimates from a wrist-worn device has been investigated. The respiratory rate was successfully derived in all cases using $m[n]$ obtained from the PVP signal, outperforming the $d_{\text{STT}}[n]$ from PVP and $m[n]$ and $d_{\text{STT}}[n]$ from PPG. P_{HF} from PVP in a bandwidth dependent on respiratory rate estimates showed a significant decrease with increasing respiratory rate, supporting its use as a parasympathetic estimate.

Bibliography

- [1] David Hernando, Reid McCallister, Jesus Lazaro, Kyle Hocking, Eduardo Gil, Bret Alvis, Pablo Laguna, Colleen Brophy, and Raquel Bailón. Validity of Venous Waveform Signal for Heart Rate Variability Monitoring. In *2018 Computing in Cardiology Conference (CinC)*, pages 1–4, Maastricht, Netherlands, 12 2018. IEEE.
- [2] David Hernando, Pablo Laguna, Colleen Brophy, Raquel Bailon, Mimma Nardelli, Kyle Hocking, Jesus Lazaro, Bret Alvis, Eduardo Gil, Enzo P. Scilingo, Daniel R. Brophy, and Gaetano Valenza. Effect of yoga on pulse rate variability measured from a venous pressure waveform. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 372–375. IEEE, 7 2019.
- [3] Kyle M. Hocking, Ban Sileshi, Franz J. Baudenbacher, Richard B. Boyer, Kelly L. Kohorst, Colleen M. Brophy, and Susan S. Eagle. Peripheral Venous Waveform Analysis for Detecting Hemorrhage and Iatrogenic Volume Overload in a Porcine Model. *Shock*, 46(4):447–452, 10 2016.
- [4] Alberto Hernando, Jesús Lázaro, Eduardo Gil, Adriana Arza, Jorge Mario Garzón, Raúl López-Antón, Concepción De La Camara, Pablo Laguna, Jordi Aguiló, and Raquel Bailón. Inclusion of Respiratory Frequency Information in Heart Rate Variability Analysis for Stress Assessment. *IEEE Journal of Biomedical and Health Informatics*, 20(4):1016–1025, 7 2016.
- [5] J. A. Hirsch and B. Bishop. Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate. *American Journal of Physiology*

Heart and Circulatory Physiology, 241(4):H620–H629, 10 1981.

- [6] Jesús Lázaro, Eduardo Gil, José María Vergara, and Pablo Laguna. Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children. *IEEE Journal of Biomedical and Health Informatics*, 18(1):240–246, 2014.
- [7] Javier Mateo and Pablo Laguna. Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal. *IEEE Transactions on Biomedical Engineering*, 50(3):334–343, 3 2003.
- [8] Jesús Lázaro, Yunyoung Nam, Eduardo Gil, Pablo Laguna, and Ki H. Chon. Respiratory rate derived from smartphone-camera-acquired pulse photoplethysmographic signals. *Physiological Measurement*, 36(11):2317–2333, 10 2015.

Effects of missing data in HRV metrics

1 Objectives and motivations

The analyses presented in the preceding chapter are limited by the absence of an evaluation regarding the degradation of metrics resulting from data loss and delineation failures. This raises several critical questions: to what extent can a given metric degrade with data loss, which metrics demonstrate greater robustness, which types of data loss are more detrimental, and, most importantly, what methodologies can be employed to conduct a more robust analysis.

In the present study, it is shown how data loss affects HRV metrics in the time domain, frequency domain, and Poincaré plots. A gap-filling method is proposed and compared with other existing approaches to alleviate these effects, both in simulated and real missing data. Two different data loss scenarios have been simulated: i) scattered missing beats, related to a low signal-to-noise ratio; and ii) bursts of missing beats, the most common due to motion artifacts. In addition, a real database of photoplethysmography-derived pulse detection series provided by the Apple Watch during a protocol that includes the relax and stress stages is analyzed. The best correction method and maximum acceptable missing beats are given. The findings of this work are useful for the design of robust HRV applications depending on the missing data tolerance and the desired HRV metrics.

2 Introduction

The exponential growth of wearable devices capable of recording ECG and/or PPG signals opened a new horizon for HRV, allowing massive monitoring at a relatively low cost. The accessibility to a large variety of designs has made them an everyday-use tool, allowing noninvasive health monitoring in the general population. In this context, assessing the state of the ANS during daily life has become a very attractive objective in the field of health and well-being. However, obtaining reliable variability measurements from wearable devices is a challenge. Wearable devices are worn throughout the day in constantly changing conditions, and motion artifacts are very frequent. In addition, comfortability is relevant when deciding the place of recording of a wearable device, in contrast to the clinical settings, where the signal quality is usually more relevant. All of this leads to an overall low signal quality compared to clinical monitoring scenarios, which downgrades the performance of traditional HRV methods. Most devices only measure MHR, which is very robust to data loss in stationary conditions but less powerful for ANS assessment than HRV. Although changes in MHR are mainly induced by the ANS, it cannot be considered a measure of autonomic function [1–3]. Despite studies that criticize the added value of HRV with respect to MHR [4], there are scenarios in which an alteration of the ANS function produces changes in HRV but not in MHR, such as in depressed patients with respect to controls [5] or in exercise contexts [6].

Acquisition technology has made a qualitative leap that has surpassed traditional HRV preprocessing methods to some extent. In a few years, the challenge has shifted from dealing with casual artifacts to being forced to forego a large part of the total recording time. The proliferation of wearable device health applications makes it necessary to investigate HRV metrics degradation in the presence of incomplete recordings, as well as new methods that allow robust analysis under adverse conditions.

2.1 Related work

Artifacts have been a concern since the beginning of HRV studies, as they can appear even in the most controlled environments. Most of the works in the literature focus on artifacts of short duration, which are often treated in the same way as ectopic beats [7–13]. In general, methods are divided between those that simply remove outliers in beat detections —both false positives and false negatives— and those that interpolate them based on accepted proximal values (gap-filling methods) [9]. Correction methods are mandatory since errors representing less than 0.1% of the detections may cause variations up to 50% in some HRV metrics [9].

Some gap-filling methods generate evenly spaced interpolations. The beat event series is not available with these methods, so time-domain metrics or Poincaré plots cannot be assessed. Mateo and Laguna proposed an IPFM-model based corrector for ectopic beats on the heart timing signal [7], a continuous signal, assuming that autonomic modulation can be modeled using a band-limited signal. Meanwhile, McNames *et al.* used an impulse rejection filter on the instantaneous heart rate signal —evenly sampled—, on the basis that nonpathological artifacts are of small duration and large amplitude [8]. Lee and Yu detected and corrected outliers in the tachogram using cubic splines [10].

On the other hand, some studies obtain a corrected unevenly-sampled inter-beat interval (IBI) series, allowing the assessment of time-domain metrics and Poincaré plots. Begum *et al.* used k-nearest neighbors in the IBI series [11], while Al Osman *et al.* used a combination of cubic and nonlinear predictive interpolation methods [12]. An interesting aspect of the last is the use of simulation to introduce artifacts in order to compare errors. Giles and Draper compared different interpolation methods of the IBI series, including cubic splines [13].

Although the previous methods may work for isolated outliers, they have not been evaluated for longer artifact segments. Baek and Shin studied the degradation of temporal and frequency metrics in response to an increase in missing IBI data, obtained by simulation, although they do not provide any

correction method [14]. The simulation randomly removes samples from the tachogram in an increasing manner, over a fairly wide range, from 5 to 285 samples, in 5-minute recordings. Morelli *et al.* developed one of the first studies to investigate the effect of a large number of heartbeat losses, from the perspective of wearable devices [15]. Their simulation method for missing detections is based on a two-state Markov chain, simulating losses of 30, 50, and 70% of IBIs. This is one of the most comprehensive studies on artifact correction applied to wearables, including temporal and frequency metrics and Poincaré plots. Benchekroun *et al.* used filtering and gap filling using a Gaussian distribution in IBI series with 5 to 35% simulated missing beats [16]. HRV metrics were derived from corrected series and used as features for a stress/relax classification. Classification results were compared with other gap-filling approaches (linear, spline, and pchip). Nevertheless, no separate metric results were reported. Królak *et al.* proposed a gap-filling algorithm tested with bursts of up to 7 missing beats [17]. They reported that cubic interpolation can, in some cases, result in lower errors for long gaps. Finally, some works address artifact correction in the detection stage, using methods such as adaptive filtering, wavelet transform, or feature extraction of the cardiac signal [18, 19]. These approaches are beyond the scope of this thesis, as they are signal specific and many wearables do not allow exporting cardiac signals but event series. In addition, they can be used in conjunction with event series correction.

2.2 Objectives and motivations

There is still much to be known about HRV metrics degradation in scenarios with large missing data. To the authors' knowledge, there is no study that provides insight into how correction methods behave under different types of loss that can occur in a real case: bursts and scattered missing beats. There is also no conclusion on the maximum burst size to discard a segment from further analysis. The same is true for scattered missing beats. In this work, degradation of different HRV metrics in the time domain, frequency domain, and Poincaré plots is evaluated in missing data scenarios. A missing data

simulation protocol has been developed for this purpose. Also, a method to attenuate the effect of missing data in HRV metrics has been proposed and compared to existing methods in the literature. Then, these methods have been applied to analyze PRV derived from Apple Watch. This work aims to contribute to HRV/PRV analysis by proposing guidelines to select the best correction method for each studied metric and missing data scenario, and to provide conclusions about when to discard a segment for further analysis depending on the quantity and distribution of missing data.

3 Materials and Methods

3.1 Simulation of missing beats

The simulation study is based on a database previously collected for a different investigation [20] comprising 16 subjects (age 28.5 ± 2.8 years, 10 men) who underwent a tilt table test consisting of 4 min in a supine position, 5 min at a 70° angle, and 4 min back to the supine position. The tilt table test is a common method used in the clinic to assess the cause of unexplained fainting. During the test, the patient's heart is monitored by electrodes while it is held in a supine position. The patient rests on a special table that has an automatic mechanism to eventually move to an upright position. This table usually has a footboard and straps to hold the patient. The purpose of the test is to trigger symptoms that indicate abnormal control of heart rate or blood pressure, which is necessary in the transition from the supine position to the upright position. It is assumed that the slight tachycardia that usually accompanies the upright position is due to a sympathovagal shift towards a sympathetic predominance in the neural modulation of the SA node. Enhanced sympathetic drive to the heart is associated with a marked increase in LF and with a decrease in the variability and arterial pressure component of HF [21]. Moreover, the grade of sympathovagal shift is strongly correlated with the degree of tilt, as demonstrated in [22]. Due to its properties, the tilt table test has become a popular method to evaluate sympathovagal balance.

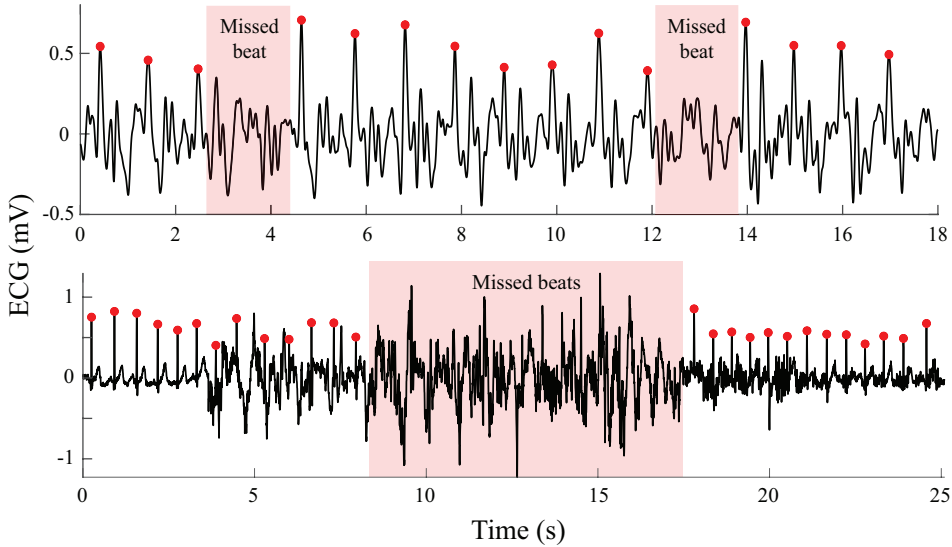


Figure 25: ECG with missed beat detections due to noise and artifacts. R-wave detections are marked with red circles. *Top*: Scattered missing beats due to a low SNR. *Bottom*: Burst of missing beats caused by an artifact.

Passive tilt only entails minimal involvement of central drive and muscular activity and is compatible with accurate stationary conditions, different from physical exercise, which is characterized by extreme reductions in HRV, non-linearities, non-stationarities, and enhanced respiratory activity [22].

A V4 lead ECG signal was recorded using the Biopac ECG100C amplifier and disposable Ag–AgCl electrodes with a sampling frequency of 1000 Hz. Two 2-minute duration segments, free of artifacts and ectopic beats, are selected for each subject, one for the first supine stage and the other for the tilt stage. Stationarity is assumed for this duration [20]. A wavelet-based algorithm has been used for QRS detection [23]. Detections were visually inspected and corrected if necessary.

The assessment of HRV metric degradation focuses on its error and its effectiveness in differentiating between tilt and supine positions, which reflect varying levels of sympathovagal balance. First, HRV metrics are computed prior to data removal, resulting in a benchmark for each method under review. Then, missing beats are simulated by removing detections from the time series in two ways: 1) by a random selection using a binomial

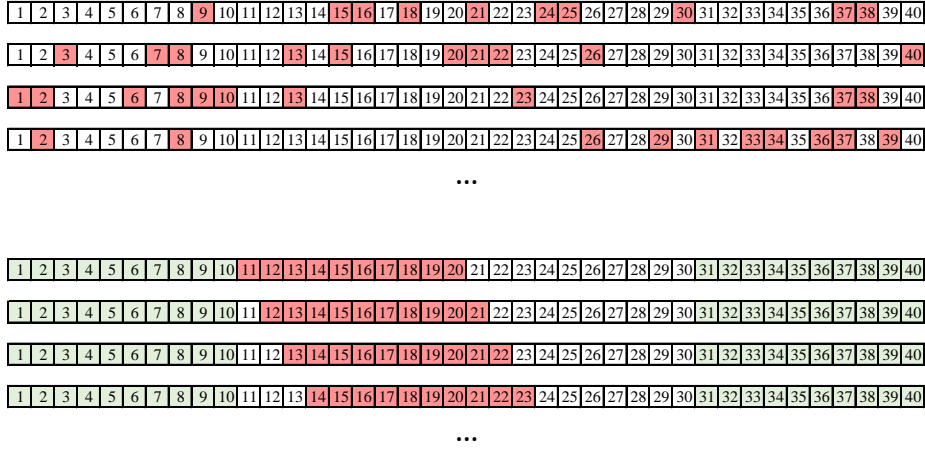


Figure 26: Example of simulation with a segment of 40 beats. Deleted beats are displayed in red. *Top*: Random distributed missing beats, $p = 0.25$. *Bottom*: Bursts of missing beats. The elements at the ends (green) cannot be deleted.

distribution; and 2) through deletion bursts, with an increasing number of missing beats in each one. The former simulates the effect of a low signal-to-noise ratio (SNR). Sometimes, signals have enough quality to perform detections, although an automatic detector can still miss some pulses in borderline situations (see Figure 25, *Top*). A binomial distribution is used to simulate this effect, so each beat is deleted with a p probability. Thus, every beat deletion is an independent Bernoulli trial. Ten different realizations of this stochastic process have been computed for each segment, obtaining a total of 160 segments for each supine and tilt position. Figure 26, *Top* shows an example of a 40-beat segment, in which 25% of the samples are removed ($p = 0.25$). In successive realizations, the positions of the removed samples change randomly.

However, artifacts can affect signals even with a high long-term SNR. Movements are mainly the cause of this kind of noise, a common problem in wearables, characterized by a finite duration and a total masking of the physiological signal. These events cause a burst of missing detections (see Figure 25, *Bottom*). This effect has been simulated by removing central elements from the series with windows of a certain duration. Although

bursts are possible at any position by taking random segments of a signal, bursts have not been simulated at interval ends, since the most advisable solution in that case is not to use those first or last seconds of the window. Specifically, 30 seconds at each of the two ends of the segment are not considered for removal. Beat removal is restricted to the remaining segment. The samples are removed from the segments with a sliding window of 10 steps, again obtaining 160 segments per supine/tilt position. An example is shown in Figure 26, *Bottom*. As the duration of the bursts is determined in seconds, a different number of beats is removed at each step, depending on the instantaneous heart rate, even for the same segment. For simplicity, in Figure 26 all bursts have the same number of elements.

In scenarios with scattered missed beats, an increase in missed beats poses a challenge in detecting where each missed beat is located, as the baseline can be lost. However, if a correct detection has been made, correction is still straightforward as adjacent beats are present. On the other hand, in the case of bursts of missing beats, detection becomes easier as the greater the number of missing beats, as they produce a larger outlier. In this case, the complication lies in finding out how many beats are missing and how to perform corrections based on gap-filling methods.

In this work, no scenarios with possible extra detections, *i.e.* false positives, are analyzed. This decision is made on the basis that only a few false positives would complicate any correction due to loss of reference. Therefore, it is assumed that detections should be performed after a signal quality evaluation stage sufficiently restrictive to avoid most false positives.

3.2 Apple Watch dataset

As a real case, the dataset described in Hernando *et al.* in [24] has been selected. It is composed of 20 healthy subjects (age 31.3 ± 8.2 years, 12 males) while undergoing a protocol that involves controlled environments of relaxation and stress. Three segments of two minutes per subject, the same duration as the simulation, have been used for each relaxation and stress phase, yielding a total of 120 segments. Two series related to heart rate

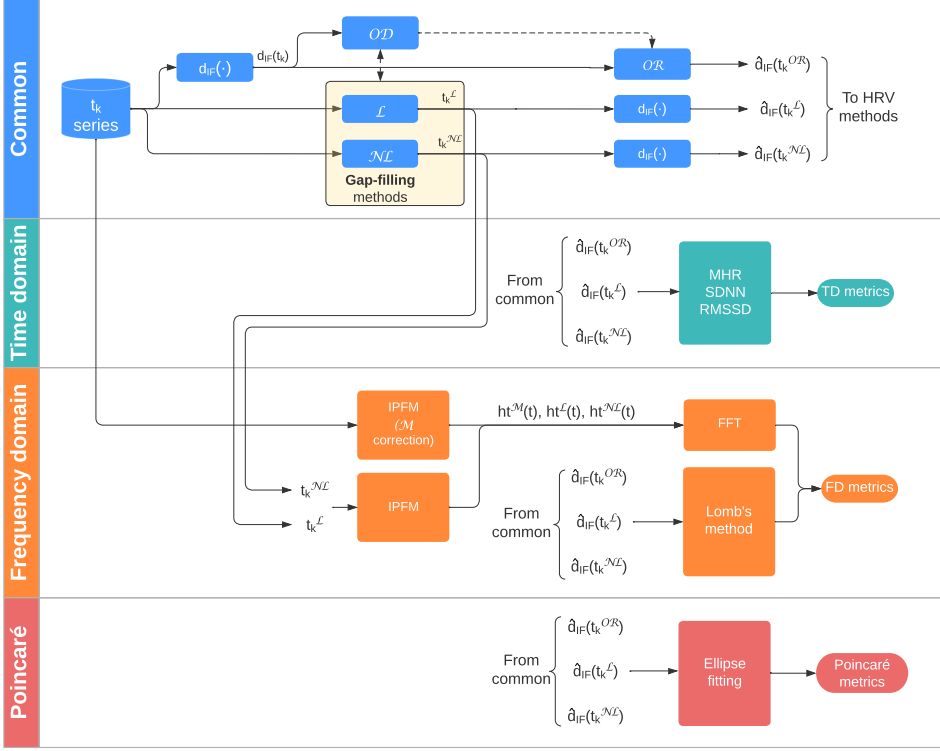


Figure 27: Process flow. The first row describes common event-series processing steps, including outlier detection and rejection, as well as gap-filling methods. After these common steps, the different HRV metrics, grouped by families, are presented in each row. OD = Outlier Detection; OR = Outlier Rejection; \mathcal{L} = Linear; \mathcal{NL} = Non-Linear; \mathcal{M} = Model-based.

were obtained in each segment: PPG-based pulse detection series recorded by the Apple Watch on the wrist and ECG-based R-wave detection series recorded by Polar H7 (Polar Electro Ltd, Kempele, Finland), the latter used as a benchmark. It is worth noting that the Apple Watch outputs the event timestamps only when the —internal— PPG allows reliable pulse detection according to an internal signal quality algorithm. Thus, the derived pulse-to-pulse series present intermittent gaps. A total of 206 gaps were found in the recordings, equivalent to 1321 missing detections. Missing data represent around 10% of total events, distributed in 6s-long gaps on average. The minimum gap length is 3.3s and the maximum is 10.4s. The synchronization

between the Apple Watch and Polar H7 was performed using the delay that maximized cross-correlation using the first 20 samples, where no gaps appeared in the Apple Watch recordings [24].

3.3 Missing data detection

Figure 27 displays a graphical summary of the methods applied, described in Sections 3.3, 3.4 and 3.5. The interval series are represented using the interval function $d_{\text{IF}}(t_k)$ (Equation 4.10). Missing beats produce outliers in $d_{\text{IF}}(t_k)$ at each t_k corresponding to events after a gap. A moving median threshold is used as an outlier-detection (\mathcal{OD}) rule, as explained in Chapter 4, Section 5. The values of α and L have been empirically set using the simulation dataset, resulting in $\alpha = 1.5$ and $L = 25$. The best value for α has been searched between 1 and 1.7 with a step of 0.1. Similarly, the best value for L has been searched between 5 and 50 with a step of 5.

3.4 Correction methods

The simplest correction rule is to remove outliers from $d_{\text{IF}}(t_k)$. This method is referred to as Outlier Removal (\mathcal{OR}) in this work, and its estimates are denoted $t_k^{\mathcal{OR}}$. However, some metrics are greatly affected by incomplete interval series. Thus, methods for estimating missing beat locations remain very interesting.

A novel gap-filling method is proposed as follows. First, missing beats are estimated by interpolation, allowing a single beat per gap. The outlier-detection rule (Equation 4.12) is applied to each new estimate, setting $\alpha = 1.1$ for a better fit. If a gap is still detected, the algorithm discards the added beats and passes to the next gap. In the next iteration, it will try to fill it with one more beat. Otherwise, it is checked if $d_{\text{IF}}(t_k) > (\beta \times d_{\text{EIBI}}(t_k))$ for all the added t_k , being $\beta = 0.9$, to avoid introducing more beats than necessary. If this condition is not fulfilled, the gap is filled with the number of beats from the previous iteration and marked as corrected. Both α and β have been empirically set using the simulation dataset. The best value for α

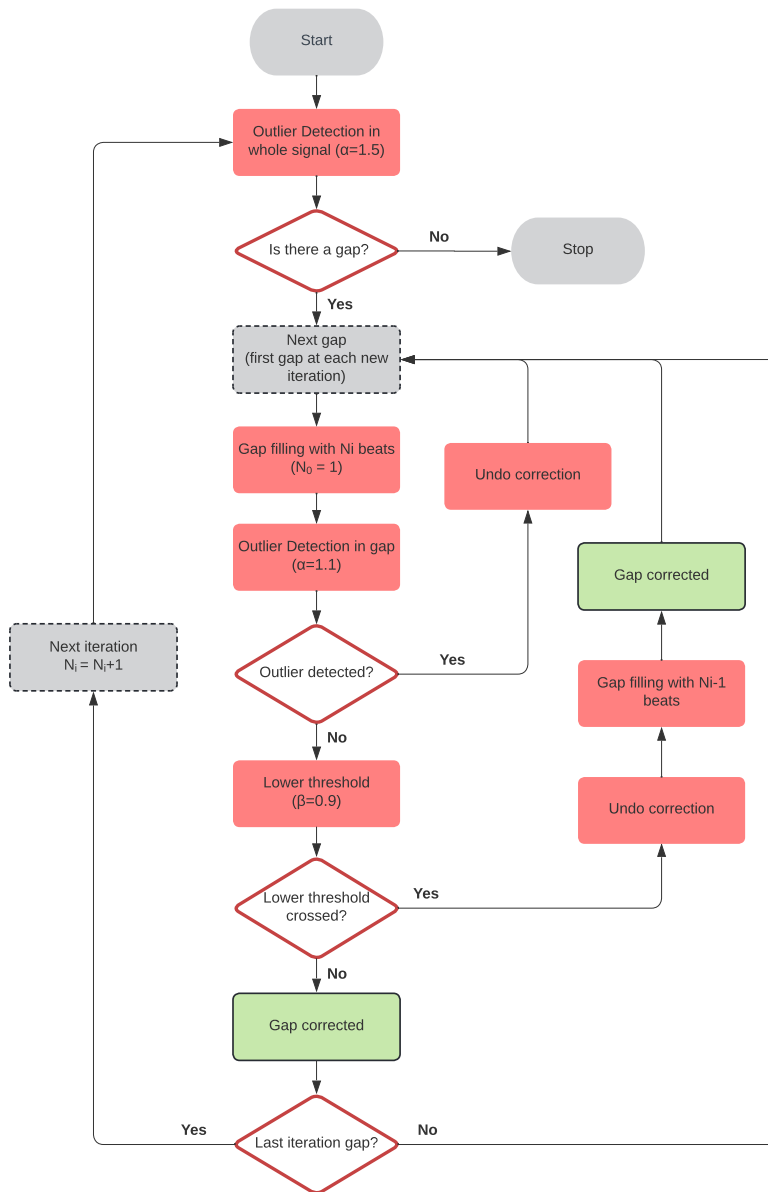


Figure 28: Gap-filling algorithm flowchart. Gray ovals represent terminal points, red rectangles represent processes, gray rectangles represent iterations, diamonds represent decisions, and green rectangles represent success in each gap correction.

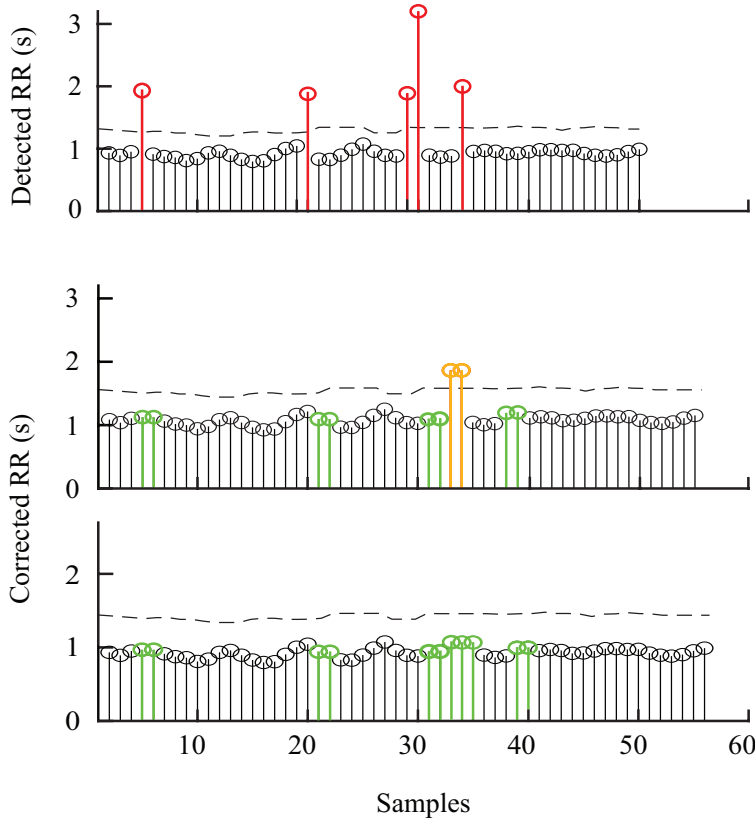


Figure 29: Demonstration of the \mathcal{L} gap-filling method. The method prioritizes filling small gaps first. *Top*: Detected R-R intervals. The dashed line represents the moving median threshold, with intervals above it (in red) requiring modification. *Center*: End of the first iteration. Successfully corrected gaps are shown in green. An uncorrected gap (in orange) remains due to an insufficient number of insertions. This gap will be addressed in subsequent iterations. *Bottom*: End of the second iteration. The gap at sample 33 is now correctly filled with the appropriate number of insertions.

has been searched between 1 and 1.5 with a step of 0.1, while the best value for β has been searched between 0.5 and 1 with a step of 0.1. At the end of the iteration, *i.e.*, when all the gaps have been covered, the outlier-detection rule is checked again in the whole segment. If not passed, a new iteration is started, using one more beat per gap until the segment is completed. A flowchart of this algorithm is presented in Figure 28.

The interpolation method greatly affects the results. Here, both linear

interpolation and nonlinear interpolation with Hermite polynomials have been used. Hermite polynomials preserve data shape and have already been shown to outperform other methods in HRV gap-filling applications [25]. In the following, gap-filling methods are referred to as linear (\mathcal{L}) and nonlinear (\mathcal{NL}) gap filling, and their estimations $t_k^{\mathcal{L}}$ and $t_k^{\mathcal{NL}}$, respectively. An example of the \mathcal{L} gap-filling method is provided in Figure 29.

Finally, the correction method described by Mateo and Laguna in [7], has also been used when analyzing metrics in the frequency domain using Fourier-based techniques. This method is referred to as Model-based (\mathcal{M}) correction. \mathcal{OR} , \mathcal{L} , \mathcal{NL} and \mathcal{M} corrections are tested both in scattered missing beat and burst simulations.

3.5 HRV metrics

Metrics in the time, frequency, and Poincaré-related domains have been computed:

- *Time domain*: Mean Heart Rate (MHR), Standard Deviation of the Normal-to-Normal interval (SDNN) and Root Mean Square of Successive Differences (RMSSD).
- *Frequency domain*: LF and HF powers (P_{LF} , P_{HF}); LF power measured in normalized units (P_{LFn}); and $P_{\text{LF}}/P_{\text{HF}}$ ratio. Only relative errors of P_{LF} and P_{HF} are presented, as the other two are derived from them. Although all subjects are included when measuring relative errors, not all could be included when measuring the ability to distinguish sympathovagal balance. For this comparison, only subjects with respiratory rates above the classic LF band (> 0.15 Hz) have been selected, thus allowing a correct frequency component separation [26]. Therefore, the simulation dataset is reduced from 16 to 9 subjects (age 28.3 ± 2.6 years, 5 males). This selection only applies when comparing metrics in the frequency domain. No selection is made in the Apple Watch dataset. In addition, the respiratory rate does not exceed 0.4 Hz, the classic HF band upper limit, in any case.

Spectral estimation is performed via Fourier-based spectral estimation and averaged Lomb-Scargle periodograms (see Chapter 4, Section 6.2). Welch's method is applied using Hamming 60-second windows, with a 50% overlap. For 120-second signals, 3 periodograms are averaged. Powers are computed using trapezoidal integration and classic windows (0.04-0.15 Hz for LF and 0.15-0.4 Hz for HF). This Fourier-based approach has been tested using model-based correction (\mathcal{M}) and gap-filling correction (\mathcal{L} and \mathcal{NL}).

On the other hand, Lomb-Scargle periodograms can be computed from unevenly spaced signals, even in the presence of missing beats. Therefore, this method has been tested both using \mathcal{OR} and gap-filling correction. Lomb-Scargle periodograms are averaged using 60-second Hamming windows with 50% overlap and powers are computed using trapezoidal integration within the classic windows as well.

- *Poincaré plots*: $SD1$, $SD2$, $SD1/SD2$, ellipse area ($S = \pi \cdot SD1 \cdot SD2$), mean distance to the ellipse centroid (Md), and standard deviation to the ellipse centroid (Sd) have been computed using the ellipse fitting method. As S and $SD1/SD2$ are computed from $SD1$ and $SD2$, relative errors are not shown for these metrics.

3.6 Statistical analysis

Relative errors (ϵ) have been calculated as the absolute value of the difference between the reference and the correction divided by the reference value, both in the simulation study and in the real database. Values are expressed as a percentage. In the simulation case, ϵ is obtained for each correction method and within each method for each type and number of removed beats. In the Apple Watch case, only one ϵ is shown for each method, since missing beats are given by the dataset (Section 3.2). ϵ is presented as a tuple of three elements: *median (first quartile - third quartile)*. A Wilcoxon signed-rank test was performed to compare the methods' performance on the same segments.

On the other hand, another signed rank test has been applied for the ANS state discrimination results. The test is done to supine/relax and tilt/stress records as separate samples, pairing states from the same subject. Metrics that could not differentiate states in any case have been omitted. Coverage graphs are also shown for the Apple Watch dataset. These graphs show the percentage of cases (n_{th}) with a relative error below a certain threshold (ϵ_{th}) as ϵ_{th} increases. These results can be very valuable for choosing a correction method based on the tolerance allowed for each application. Coverage graphs are not included for the simulation due to the large number of combinations depending on the type and number of deletions. Segment rejection decision thresholds, *i.e.*, the maximum deletion probability/burst duration allowed to obtain reliable results are also proposed in Section 5. These thresholds are proposed based on the criterion that the third quartile of the relative error does not exceed 20%.

4 Results

4.1 Time-domain metrics

Table 3 shows relative error values of the different metrics with increasing deletion probability, in the case of scattered missing beats (Table 3a) and burst duration (Table 3b). Regarding the relative error of scattered missing beats, \mathcal{NL} gap filling is the best performing correction method for MHR and SDNN for all deletion probabilities, although no significant differences can be found between \mathcal{OR} and \mathcal{NL} up to 35% missing beats in the case of MHR. \mathcal{L} gap filling yields the best results for RMSSD up to 25% deletion probability. A higher degradation can be observed at high loss rates, being \mathcal{OR} the best option from 25% deletion probability onward. In the case of bursts, \mathcal{NL} gap filling yields the best results for MHR up to 10 s bursts. No significant differences can be found between \mathcal{OR} and \mathcal{NL} from 15 s. \mathcal{OR} gives the best results for SDNN and RMSSD.

Table 3: Relative error (%) of time-domain metrics. (a) Scattered missing beats. **(b)** Bursts. †: Significant differences ($p < 0.05$) between \mathcal{OR} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{OR} .

Method	Metric	Deletion probability (%)			
		5	15	25	35
\mathcal{OR}	MHR	0 (0–0)	0 (0–1) [†]	0 (0–1) [†]	1 (0–1) [†]
	SDNN	2 (1–3) [†]	4 (2–6) [†]	5 (2–9) [†]	7 (3–15) [†]
	RMSSD	2 (1–4) [†]	5 (2–9) [†]	9 (4–15)	11 (5–24)
\mathcal{L}	MHR	0 (0–0) ^{Δ}	0 (0–0) ^{Δ}	0 (0–0) ^{Δ}	0 (0–1) ^{Δ}
	SDNN	0 (0–1) ^{Δ}	2 (1–3) ^{Δ}	5 (3–8) ^{Δ}	8 (4–15) ^{Δ}
	RMSSD	1 (0–2) ^{Δ}	3 (1–6) ^{Δ}	8 (3–20) ^{Δ}	14 (5–38) ^{Δ}
\mathcal{NL}	MHR	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–1) [§]
	SDNN	0 (0–0) [§]	1 (0–2) [§]	3 (1–6) [§]	5 (2–10) [§]
	RMSSD	1 (0–2) [§]	4 (2–7) [§]	9 (5–17) [§]	15 (8–29) [§]

(a)

Method	Metric	Burst duration (s)			
		5	10	15	20
\mathcal{OR}	MHR	0 (0–0)	0 (0–0) [†]	0 (0–1) [†]	0 (0–1) [†]
	SDNN	2 (1–2) [†]	2 (1–4) [†]	3 (1–5) [†]	4 (2–6) [†]
	RMSSD	2 (1–2) [†]	2 (1–4) [†]	3 (2–5) [†]	4 (2–7) [†]
\mathcal{L}	MHR	0 (0–0) ^{Δ}	0 (0–1) ^{Δ}	1 (0–1) ^{Δ}	1 (0–1) ^{Δ}
	SDNN	1 (1–3) ^{Δ}	3 (2–5) ^{Δ}	5 (3–8) ^{Δ}	6 (3–10) ^{Δ}
	RMSSD	2 (1–3) ^{Δ}	4 (2–6) ^{Δ}	5 (3–8) ^{Δ}	7 (4–11) ^{Δ}
\mathcal{NL}	MHR	0 (0–0) [§]	0 (0–1) [§]	0 (0–1)	1 (0–1)
	SDNN	1 (0–3) [§]	3 (1–5) [§]	5 (2–7) [§]	6 (3–9) [§]
	RMSSD	2 (1–4) [§]	4 (2–7) [§]	6 (3–9) [§]	8 (4–12) [§]

(b)

Table 4: p -values of ranked signed test for supine/tilt discrimination of time-domain metrics. N.S.: Not Significant ($p > 0.05$).

Method	Metric	Reference	Deletion probability (%)			
			5	15	25	35
\mathcal{OR}	MHR	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	SDNN	0.020	0.011	0.016	N.S.	N.S.
	RMSSD	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
\mathcal{L}	MHR	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	SDNN	0.020	0.014	0.008	N.S.	N.S.
	RMSSD	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
\mathcal{NL}	MHR	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	SDNN	0.020	0.012	0.011	0.021	0.014
	RMSSD	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$

(a)

Method	Metric	Reference	Burst duration (s)			
			5	10	15	20
\mathcal{OR}	MHR	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	SDNN	0.020	0.011	0.011	0.011	0.011
	RMSSD	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
\mathcal{L}	MHR	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	SDNN	0.020	0.011	0.011	0.011	0.004
	RMSSD	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
\mathcal{NL}	MHR	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	SDNN	0.020	0.014	0.007	0.012	0.026
	RMSSD	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$

(b)

Table 4 shows the results of the Wilcoxon test for distinguishing between supine and tilt states. The first column shows the reference test results. There are no major differences between methods, although only \mathcal{NL} is able to maintain the benchmark results throughout the entire simulation. This is consistent with \mathcal{NL} results in terms of relative error. Both \mathcal{OR} and \mathcal{L} fail with SDNN in the case of scattered missing beats with deletion probabilities greater than 15%.

Table 5 shows the relative errors in the Apple Watch dataset, exhibiting equality among all correction methods. Figure 30 shows the coverage from the Apple Watch dataset. No differences are found between methods. MHR once again demonstrates great robustness, having n_{th} close to 100% with less than 2% ϵ_{th} . SDNN achieves 80% n_{th} with 10% ϵ_{th} , while for the same ϵ_{th} , RMSSD has 60% n_{th} .

Finally, Figure 31 shows the metric distributions with relax (green) and stress (blue) groups separately from the Apple Watch dataset. Wilcoxon test results are marked with asterisks above each pair. One asterisk indicates $p < 0.05$ and two asterisks indicate $p < 0.001$. All correction methods present the same behavior for MHR and SDNN. RMSSD results do show improved \mathcal{OR} performance by maintaining the reference $p < 0.001$ versus $p < 0.05$ of the gap-filling methods.

Table 5: Relative error (%) of time-domain metrics from Apple Watch dataset. †: Significant differences ($p < 0.05$) between \mathcal{OR} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{OR} .

Metric	Method		
	\mathcal{OR}	\mathcal{L}	\mathcal{NL}
MHR	0.1 (0.0–0.5)	0.0 (0.0–0.5) Δ	0.0 (0.0–0.7) \S
SDNN	3.4 (2.0–7.5) \dagger	2.9 (1.5–9.6) Δ	3.0 (1.3–8.6)
RMSSD	7.8 (4.3–15.9) \dagger	8.6 (4.0–20.2) Δ	8.6 (3.7–17.7) \S

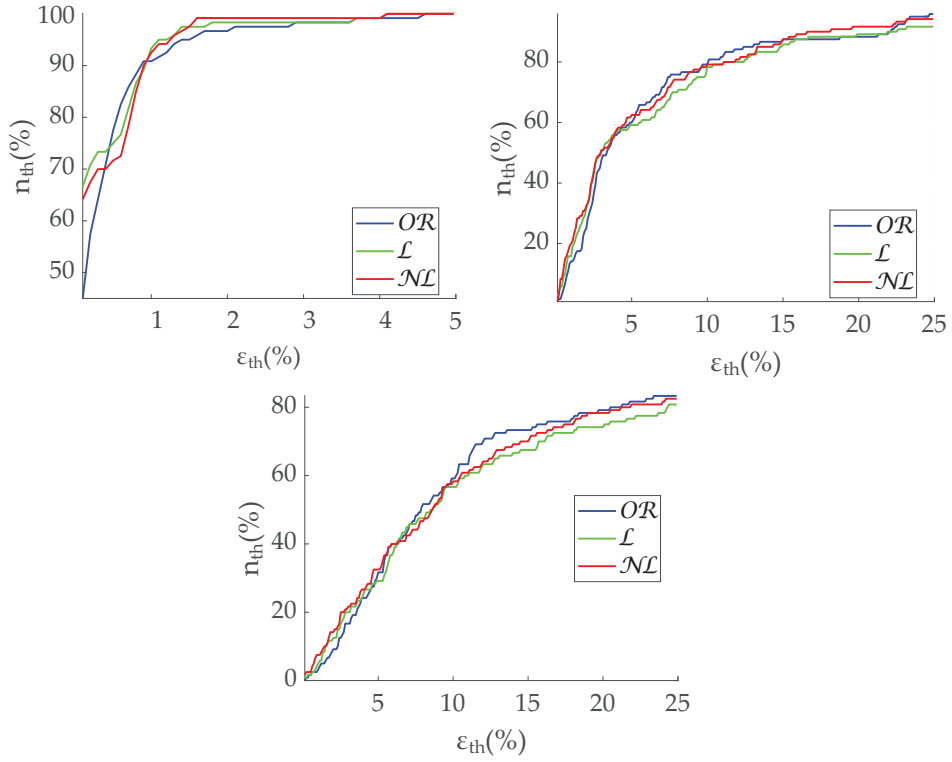


Figure 30: Coverage of time-domain metrics from Apple Watch dataset. The graphs show the percentage of cases (n_{th}) with a relative error below a certain threshold (ϵ_{th}). *Top left: MHR. Top right: SDNN. Bottom: RMSSD.*

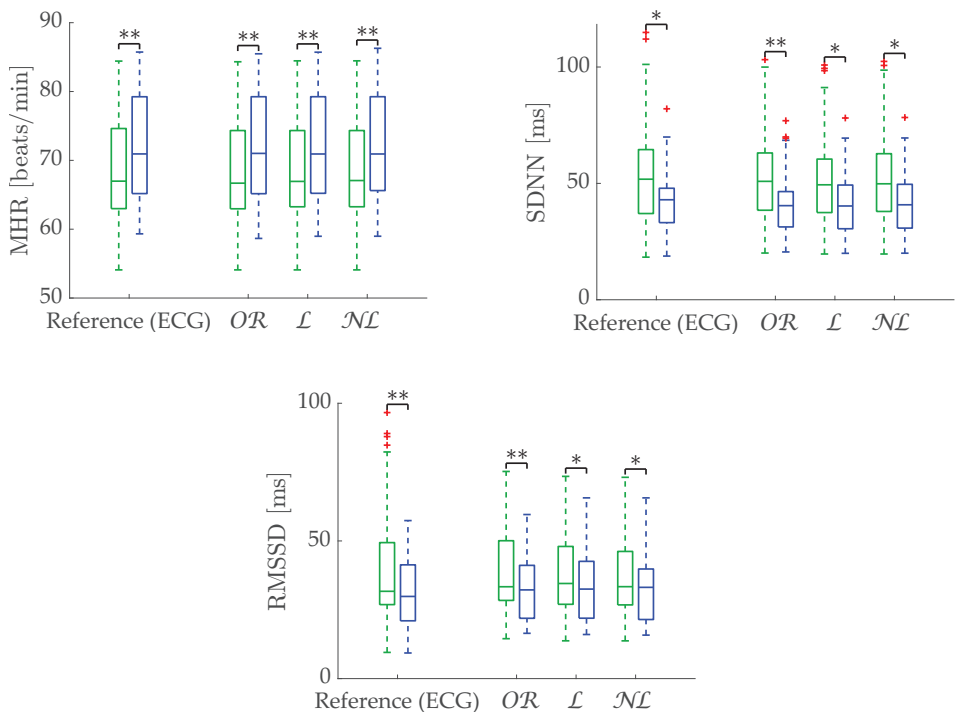


Figure 31: Relax (green)/Stress (blue) discrimination of time-domain metrics from Apple Watch dataset. *Top left: MHR. Top right SDNN. Bottom: RMSSD.* *: Significant differences ($p < 0.05$) between relax and stress groups. **: Significant differences ($p < 0.001$) between relax and stress groups.

4.2 Frequency-domain metrics computed via Fourier transform

In the case of frequency-domain metrics, gap-filling methods show a clear improvement. \mathcal{NL} gap filling is the best performing method in terms of relative error in the case of scattered missing beats (Table 6).

Table 6: Relative error (%) of frequency-domain metrics computed via FFT. (a) Scattered missing beats. (b) Bursts. †: Significant differences ($p < 0.05$) between \mathcal{M} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{M} .

Method	Metric	Deletion probability (%)			
		5	15	25	35
\mathcal{M}	P_{LF}	8 (4–19) [†]	20 (11–33) [†]	36 (22–54) [†]	55 (29–162) [†]
	P_{HF}	15 (7–30) [†]	33 (20–45) [†]	50 (37–63) [†]	59 (43–73) [†]
\mathcal{L}	P_{LF}	1 (0–2) ^{Δ}	4 (2–9) ^{Δ}	11 (6–19) ^{Δ}	16 (9–29) ^{Δ}
	P_{HF}	3 (1–6) ^{Δ}	11 (6–18) ^{Δ}	23 (12–41) ^{Δ}	34 (20–61)
\mathcal{NL}	P_{LF}	0 (0–1) [§]	1 (1–5) [§]	4 (1–13) [§]	9 (2–21) [§]
	P_{HF}	2 (1–4) [§]	7 (3–15) [§]	19 (10–38) [§]	29 (17–55) [§]

(a)

Method	Metric	Burst duration (s)			
		5	10	15	20
\mathcal{M}	P_{LF}	10 (4–21) [†]	15 (7–26) [†]	22 (10–33) [†]	27 (14–39) [†]
	P_{HF}	13 (6–28) [†]	18 (9–33) [†]	23 (14–36) [†]	29 (19–43) [†]
\mathcal{L}	P_{LF}	5 (2–12) ^{Δ}	11 (5–19) ^{Δ}	16 (7–26) ^{Δ}	19 (9–31)
	P_{HF}	7 (3–11) ^{Δ}	10 (5–17) ^{Δ}	14 (7–23) ^{Δ}	18 (9–28) ^{Δ}
\mathcal{NL}	P_{LF}	5 (2–12) [§]	10 (4–18) [§]	13 (6–25) [§]	19 (9–31) [§]
	P_{HF}	7 (3–12) [§]	11 (6–18) [§]	15 (8–24) [§]	19 (11–30) [§]

(b)

Table 7: p -values of ranked signed test for supine/tilt discrimination of frequency-domain metrics computed via FFT. N.S.: Not Significant ($p > 0.05$).

Method	Metric	Reference	Deletion probability (%)			
			5	15	25	35
\mathcal{M}	P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	N.S.
	P_{LFn}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LF}/P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	0.005
\mathcal{L}	P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LFn}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LF}/P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
\mathcal{NL}	P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LFn}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LF}/P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$

(a)

Method	Metric	Reference	Burst duration (s)			
			5	10	15	20
\mathcal{M}	P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LFn}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LF}/P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
\mathcal{L}	P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LFn}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LF}/P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
\mathcal{NL}	P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LFn}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LF}/P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$

(b)

Gap-filling correction advantage is maintained in the case of bursts. Although differences are reduced, they are still significant. Also, differences between \mathcal{L} and \mathcal{NL} gap filling are reduced. In this case, \mathcal{L} gap filling performs better for P_{HF} , while \mathcal{NL} is still better for P_{LF} . Another aspect to note is that correction is not as effective in P_{HF} as in P_{LF} with scattered missing beats.

Discrimination results follow a similar pattern (Table 7). For scattered missing beats, gap-filling correction performed better than \mathcal{M} correction for P_{HF} and P_{LF}/P_{HF} . This difference only appears after a 35% deletion probability; thus, differences are not very large. On the other hand, results are identical for the burst case. P_{LF} showed no discrimination capacity for this dataset.

Regarding the Apple Watch dataset, \mathcal{NL} gap filling obtains the best performance at low frequencies (Table 8), although there is virtually no difference at high frequencies. In addition, P_{HF} errors are higher than P_{LF} errors as in the simulation.

Coverage graphs show the same phenomena (Figure 32). P_{LF} coverages are similar until 10% ϵ_{th} —approximately 60% n_{th} —, separating thereafter. \mathcal{NL} gap filling is the best correction method, followed by \mathcal{L} gap filling. In contrast, there are no differences for the P_{HF} case. Also, the coverage is clearly lower, approximately 40% n_{th} at 10% ϵ_{th} . Both P_{LF} and P_{HF} correctly discriminate the states (Figure 33), showing no difference between correction methods.

Table 8: Relative error (%) of frequency-domain metrics computed via FFT from Apple Watch dataset. †: Significant differences ($p < 0.05$) between \mathcal{M} and \mathcal{L} . \triangle : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{M} .

Metric	Method		
	\mathcal{M}	\mathcal{L}	\mathcal{NL}
P_{LF}	0.1 (0.0–0.3) [†]	0.1 (0.0–0.2) ^{\triangle}	0.1 (0.0–0.2) [§]
P_{HF}	0.1 (0.1–0.3) [†]	0.2 (0.1–0.3) ^{\triangle}	0.2 (0.1–0.3) [§]

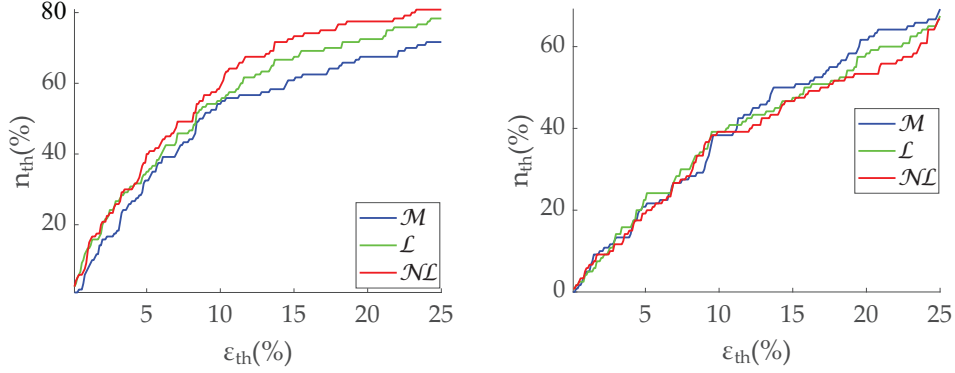


Figure 32: Coverage of frequency-domain metrics computed via Fourier transform from Apple Watch dataset. The graphs show the percentage of cases (n_{th}) with a relative error below a certain threshold (ϵ_{th}). *Left:* P_{LF} . *Right:* P_{HF} .

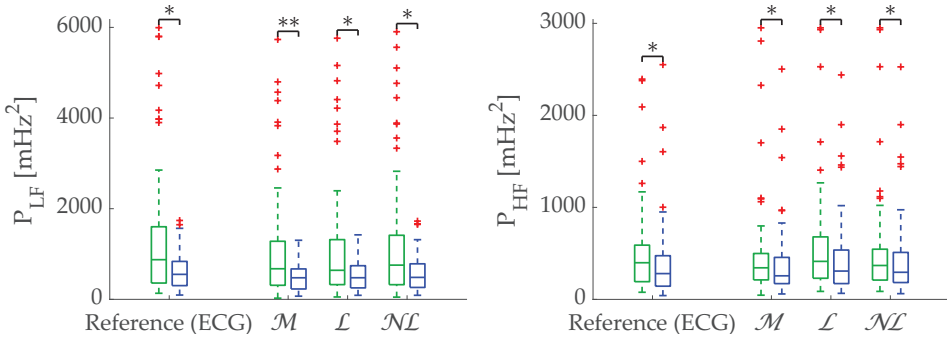


Figure 33: Relax (green)/Stress (blue) discrimination of frequency-domain metrics computed via Fourier transform from Apple Watch dataset. *Left:* P_{LF} . *Right:* P_{HF} . *: Significant differences ($p < 0.05$) between relax and stress groups. **: Significant differences ($p < 0.001$) between relax and stress groups.

4.3 Frequency-domain metrics computed via Lomb-Scargle periodograms

In the case of frequency-domain results calculated via Lomb-Scargle periodograms, \mathcal{NL} gap filling clearly outperforms the others with scattered missing beats, as well as for P_{LF} with small bursts (Table 9).

Table 9: Relative error (%) of frequency-domain metrics computed via Lomb-Scargle periodograms. (a) Scattered missing beats. (b) Bursts. †: Significant differences ($p < 0.05$) between \mathcal{OR} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{OR} .

Method	Metric	Deletion probability (%)			
		5	15	25	35
\mathcal{OR}	P_{LF}	11 (5–19) [†]	24 (10–41) [†]	35 (18–66) [†]	58 (25–123) [†]
	P_{HF}	23 (11–46) [†]	79 (47–155) [†]	160 (87–297) [†]	305 (143–665) [†]
\mathcal{L}	P_{LF}	1 (0–2) ^Δ	4 (2–8) ^Δ	10 (5–18) ^Δ	16 (8–29) ^Δ
	P_{HF}	3 (1–5) ^Δ	9 (5–16) ^Δ	21 (11–37) ^Δ	31 (17–62) ^Δ
\mathcal{NL}	P_{LF}	0 (0–1) [§]	1 (0–4) [§]	3 (1–12) [§]	8 (2–23) [§]
	P_{HF}	2 (1–3) [§]	6 (2–12) [§]	16 (8–32) [§]	28 (15–53) [§]

(a)

Method	Metric	Burst duration (s)			
		5	10	15	20
\mathcal{OR}	P_{LF}	11 (7–17) [†]	19 (11–29) [†]	25 (13–38) [†]	29 (15–49) [†]
	P_{HF}	14 (8–20) [†]	23 (13–34) [†]	31 (18–45) [†]	39 (24–61) [†]
\mathcal{L}	P_{LF}	5 (2–11) ^Δ	10 (4–19) ^Δ	14 (7–24) ^Δ	17 (7–29) ^Δ
	P_{HF}	6 (2–11) ^Δ	9 (4–18) ^Δ	13 (5–22) ^Δ	16 (7–26) ^Δ
\mathcal{NL}	P_{LF}	5 (2–12) [§]	8 (4–17) [§]	14 (6–24) [§]	18 (9–29) [§]
	P_{HF}	6 (3–12) [§]	11 (5–19) [§]	15 (8–23) [§]	19 (10–29) [§]

(b)

Table 10: p -values of ranked signed test for supine/tilt discrimination of frequency-domain metrics computed via Lomb's method. N.S.: Not Significant ($p > 0.05$).

Method	Metric	Reference	Deletion probability (%)			
			5	15	25	35
\mathcal{OR}	P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LFn}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LF}/P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
\mathcal{L}	P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LFn}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LF}/P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
\mathcal{NL}	P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LFn}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LF}/P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$

(a)

Method	Metric	Reference	Burst duration (s)			
			5	10	15	20
\mathcal{OR}	P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LFn}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LF}/P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
\mathcal{L}	P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LFn}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LF}/P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
\mathcal{NL}	P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LFn}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	P_{LF}/P_{HF}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$

(b)

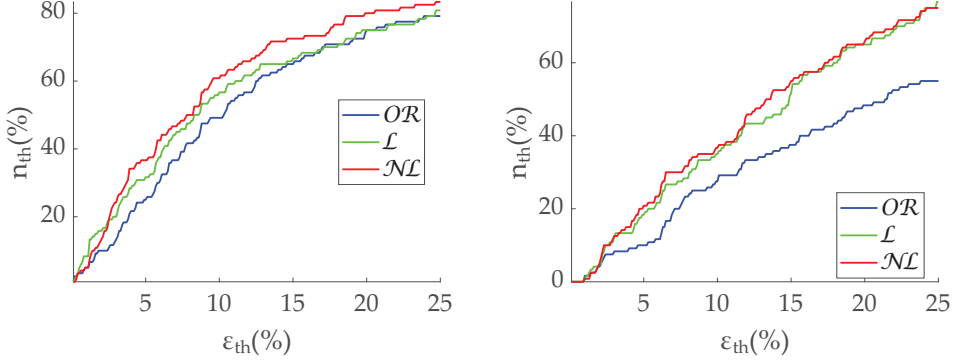


Figure 34: Coverage of frequency-domain metrics computed via Lomb-Scargle periodograms from Apple Watch dataset. The graphs show the percentage of cases (n_{th}) with a relative error below a certain threshold (ϵ_{th}). Left: P_{LF} . Right: P_{HF} .

\mathcal{L} gap filling performs better for P_{LF} from 15 seconds onwards and for P_{HF} with any burst duration. Statistically significant differences are found between all methods at any loss rate. All methods are equally reliable in terms of discrimination for all deletion probabilities and burst durations (Table 10).

\mathcal{NL} gap filling remains superior in the Apple Watch dataset in terms of relative error (Table 11), followed by \mathcal{L} gap filling. Coverage graphs (Figure 34) show an advantage of \mathcal{NL} in P_{LF} , while both \mathcal{NL} and \mathcal{L} gap filling perform similarly in P_{HF} , although much better than \mathcal{OR} . As in simulation, all methods are robust in state discrimination (Figure 35).

Table 11: Relative error (%) of frequency-domain metrics computed via Lomb-Scargle periodograms metrics from Apple Watch dataset. †: Significant differences ($p < 0.05$) between \mathcal{OR} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{OR} .

Metric	Method		
	\mathcal{OR}	\mathcal{L}	\mathcal{NL}
P_{LF}	0.1 (0.1–0.2) [†]	0.1 (0.0–0.2) ^{Δ}	0.1 (0.0–0.2) [§]
P_{HF}	0.2 (0.1–0.6) [†]	0.2 (0.1–0.3) ^{Δ}	0.1 (0.1–0.3) [§]

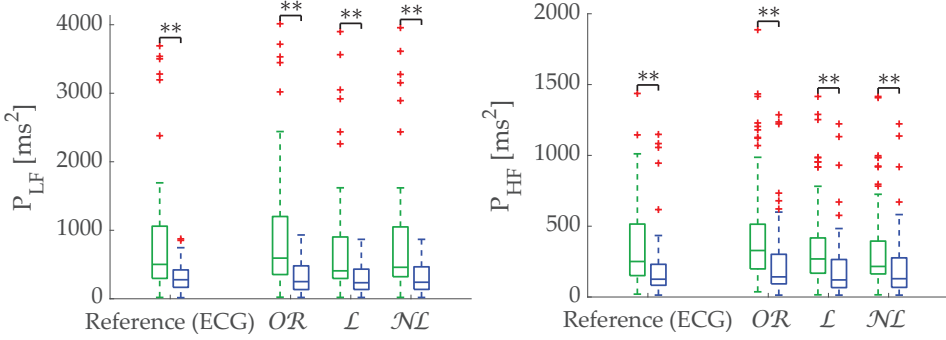


Figure 35: Relax (green)/Stress (blue) discrimination of frequency-domain metrics computed via Lomb-Scargle periodograms from Apple Watch dataset. Left: P_{LF} . Right: P_{HF} . *: Significant differences ($p < 0.05$) between relax and stress groups. **: Significant differences ($p < 0.001$) between relax and stress groups.

4.4 Poincaré Plots

As for time-domain metrics, there is no clear difference between correction methods for Poincaré metrics (Table 12).

In the case of scattered missing beats, \mathcal{L} performs better with SD1 when the deletion probability is below 25%. There are no significant differences with \mathcal{OR} from 25% onwards. \mathcal{NL} outperforms the others with SD2, Md, and Sd. On the other hand, \mathcal{OR} is the best for SD1, SD2, and Md when dealing with bursts. No significant differences can be found with Sd.

Results in terms of group discrimination suggest an advantage of \mathcal{NL} gap filling in the case of scattered missing beats, while \mathcal{NL} and \mathcal{OR} perform similarly when dealing with bursts (Table 13). The three methods perform virtually identically on the Apple Watch dataset, both in terms of relative error (Table 14); coverage (Figure 36); and discrimination (Figure 37). In the last, \mathcal{OR} performed better with SD1 and S, in coherence with the simulation.

Table 12: Relative error (%) of Poincaré metrics. (a) Scattered missing beats. **(b)** Bursts. †: Significant differences ($p < 0.05$) between \mathcal{OR} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{OR} .

Method	Metric	Deletion probability (%)			
		5	15	25	35
\mathcal{OR}	SD1	2 (1–4) [†]	5 (2–9) [†]	9 (4–14)	11 (5–24)
	SD2	3 (1–4) [†]	5 (2–9) [†]	7 (3–13) [†]	11 (5–20) [†]
	Md	2 (1–4) [†]	5 (2–7) [†]	6 (3–11) [†]	9 (4–18) [†]
	Sd	3 (1–5) [†]	6 (3–11)	10 (5–19) [†]	15 (7–31) [†]
\mathcal{L}	SD1	1 (0–2) Δ	3 (1–6) Δ	8 (3–20) Δ	14 (5–38) Δ
	SD2	0 (0–1) Δ	2 (1–3) Δ	4 (2–7) Δ	7 (4–13) Δ
	Md	1 (0–1) Δ	2 (1–4) Δ	5 (2–7) Δ	7 (4–12) Δ
	Sd	1 (0–1) Δ	1 (0–3) Δ	4 (1–13) Δ	7 (2–34) Δ
\mathcal{NL}	SD1	1 (0–2) [§]	4 (2–7) [§]	9 (5–17) [§]	14 (8–29) [§]
	SD2	0 (0–0) [§]	1 (0–2) [§]	2 (1–4) [§]	3 (1–8) [§]
	Md	0 (0–1) [§]	1 (0–3) [§]	2 (1–6) [§]	5 (2–9) [§]
	Sd	0 (0–1) [§]	1 (0–3) [§]	2 (1–8) [§]	5 (2–17) [§]

(a)

Method	Metric	Burst duration (s)			
		5	10	15	20
\mathcal{OR}	SD1	2 (1–3) [†]	3 (1–4) [†]	3 (2–5) [†]	4 (2–7) [†]
	SD2	2 (1–3) [†]	2 (1–4) [†]	3 (2–5) [†]	4 (2–6) [†]
	Md	2 (1–3) [†]	3 (1–5) [†]	4 (1–6) [†]	4 (2–7) [†]
	Sd	2 (1–3)	2 (1–4)	3 (2–5)	4 (2–6) [†]
\mathcal{L}	SD1	2 (1–3) Δ	4 (2–6) Δ	5 (3–8) Δ	7 (4–11) Δ
	SD2	1 (1–3) Δ	3 (1–5) Δ	5 (2–8) Δ	6 (3–10) Δ
	Md	2 (1–4) Δ	6 (3–9) Δ	8 (4–12) Δ	11 (5–14) Δ
	Sd	2 (1–4) Δ	3 (2–7)	4 (2–8)	5 (2–8) Δ
\mathcal{NL}	SD1	2 (1–4) [§]	4 (2–7) [§]	6 (3–9) [§]	8 (4–12) [§]
	SD2	1 (0–3) [§]	3 (1–5) [§]	4 (2–7) [§]	6 (3–9) [§]
	Md	1 (0–3) [§]	3 (1–5) [§]	4 (2–7) [§]	6 (3–9) [§]
	Sd	2 (1–4)	3 (1–6)	4 (2–8)	4 (2–7)

(b)

Table 13: p -values of ranked signed test for supine/tilt discrimination of Poincaré metrics. N.S.: Not Significant ($p > 0.05$).

Method	Metric	Reference	Deletion probability (%)			
			5	15	25	35
\mathcal{OR}	SD1	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	SD2	0.031	N.S.	N.S.	N.S.	N.S.
	SD12	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	S	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	0.012	0.002
	Md	$< 10^{-3}$	0.001	0.002	0.155	0.016
	Sd	0.039	0.009	N.S.	N.S.	N.S.
\mathcal{L}	SD1	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	SD2	0.031	N.S.	N.S.	N.S.	N.S.
	SD12	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	S	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	0.007
	Md	$< 10^{-3}$	$< 10^{-3}$	0.001	0.010	0.012
	Sd	0.039	0.034	N.S.	N.S.	N.S.
\mathcal{NL}	SD1	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	SD2	0.031	0.043	N.S.	N.S.	N.S.
	SD12	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	S	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	0.004
	Md	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	0.002	0.002
	Sd	0.039	0.033	N.S.	N.S.	N.S.

(a)

Method	Metric	Reference	Burst duration (s)			
			5	10	15	20
\mathcal{OR}	SD1	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	SD2	0.031	N.S.	N.S.	N.S.	N.S.
	SD12	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	S	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	Md	$< 10^{-3}$	0.001	0.002	0.002	0.002
	Sd	0.039	0.024	0.026	0.022	0.015
\mathcal{L}	SD1	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	SD2	0.031	N.S.	N.S.	N.S.	N.S.
	SD12	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	S	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	Md	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	0.002	0.002
	Sd	0.039	N.S.	N.S.	N.S.	N.S.
\mathcal{NL}	SD1	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	SD2	0.031	N.S.	N.S.	N.S.	N.S.
	SD12	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	S	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	Md	$< 10^{-3}$	0.002	0.001	0.002	0.013
	Sd	0.039	0.041	0.041	0.027	0.049

(b)

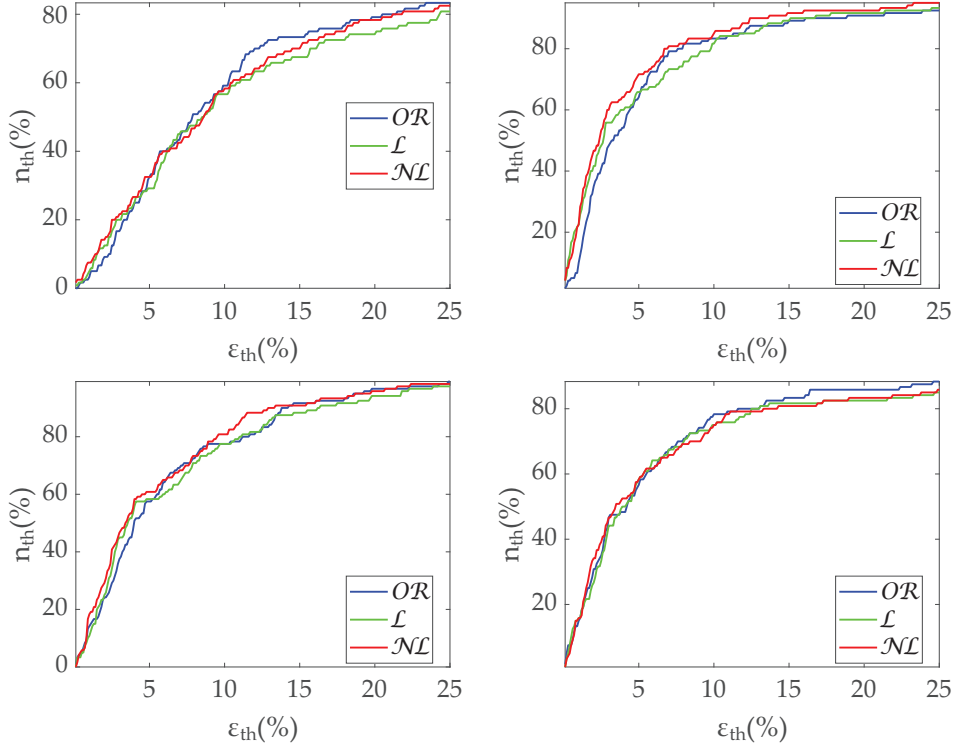


Figure 36: Coverage of Poincaré metrics from Apple Watch dataset. The graphs show the percentage of cases (n_{th}) with a relative error below a certain threshold (ϵ_{th}). *Top left: SD1. Top right: SD2. Bottom left: Md. Bottom right: Sd.*

Table 14: Relative error (%) of Poincaré metrics from Apple Watch dataset.

†: Significant differences ($p < 0.05$) between OR and \mathcal{L} . \triangle : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and OR .

Metric	Method		
	OR	\mathcal{L}	\mathcal{NL}
SD1	7.8 (4.2–15.9) [†]	8.6 (4.0–20.2) ^{\triangle}	8.6 (3.7–17.7) [§]
SD2	3.2 (1.6–6.4) [†]	2.6 (1.1–8.0) ^{\triangle}	2.4 (1.1–6.1)
Md	4.0 (2.2–8.3) [†]	3.6 (2.0–9.2) ^{\triangle}	3.4 (1.7–8.4)
Sd	4.2 (1.7–9.4)	4.0 (1.9–10.0) ^{\triangle}	3.4 (1.5–10.1) [§]

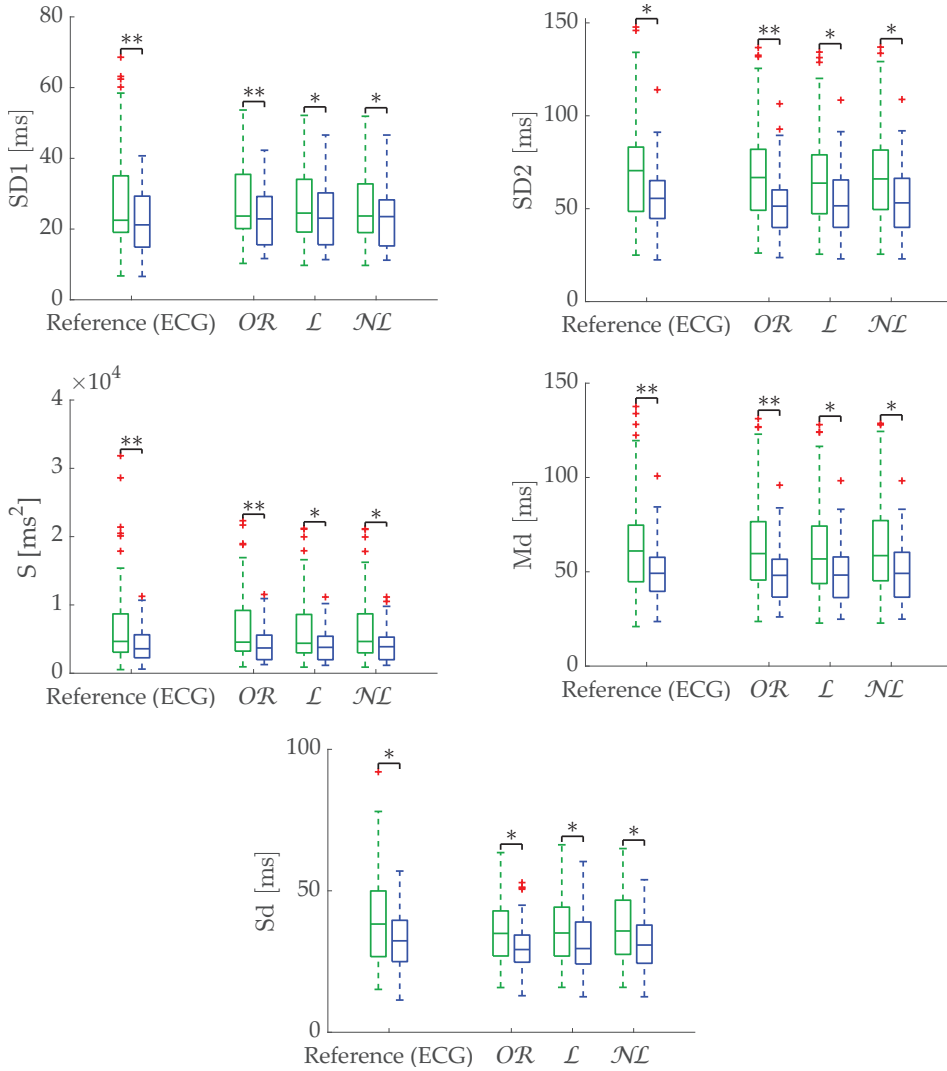


Figure 37: Relax (green)/Stress (blue) discrimination of Poincaré metrics from Apple Watch dataset. Top left: SD1. Top right: SD2. Center left: S. Center right: Md. Bottom: Sd. *: Significant differences ($p < 0.05$) between relax and stress groups. **: Significant differences ($p < 0.001$) between relax and stress groups.

5 Discussion

An analysis of the degradation of some of the most important HRV metrics due to data loss has been presented. A simulation study has been designed to test the influence of missing beats depending on whether they are scattered or in bursts. Correction methods have been tested with both simulation and experimental data, recorded with a commercial wearable. Note that, in contrast to the simulation dataset, PRV was compared to the HRV in the case of the Apple Watch dataset. Thus, the error results obtained for the simulation dataset should not be compared with those obtained for the Apple Watch dataset. Nevertheless, correction methods within the same dataset can still be compared. In the following, a discussion on the best correction method for each metric is given, as well as the maximum acceptable missing beats for a relative error of less than 20% in the third quartile. A summary is shown in Table 15.

Table 15: Summary of findings. Best correction method and maximum acceptable missing beats for a relative error less than 20% in the third quartile.

Metric	Best correction method		Max. acceptable missing beats	
	Scattered	Bursts	Scattered	Bursts
MHR	\mathcal{NL}	\mathcal{NL}	35%	20 s
SDNN	\mathcal{NL}	\mathcal{OR}	35%	20 s
RMSSD	\mathcal{L}	\mathcal{OR}	25%	20 s
P_{LF} (FFT)	\mathcal{NL}	\mathcal{NL}	25%	10 s
P_{HF} (FFT)	\mathcal{NL}	\mathcal{L}	15%	10 s
P_{LF} (Lomb)	\mathcal{NL}	\mathcal{NL}	25%	10 s
P_{HF} (Lomb)	\mathcal{NL}	\mathcal{L}	15%	10 s
SD1	\mathcal{L}	\mathcal{OR}	25%	20 s
SD2	\mathcal{NL}	\mathcal{OR}	35%	20 s
Md	\mathcal{NL}	\mathcal{OR}	35%	20 s
Sd	\mathcal{NL}	$\mathcal{OR} / \mathcal{NL}$	35%	20 s

5.1 Time-domain metrics

Regarding time-domain metrics, only noticeable differences are found in the relative error results of the simulation. \mathcal{NL} is the best option in the case of applications where MHR is the only interesting metric, as it is the best correction method both with bursts and scattered missing beats. \mathcal{NL} is also the best performing method for SDNN with scattered losses. \mathcal{OR} is a reliable correction for SDNN and RMSSD in datasets with bursts predominance, while RMSSD should be computed using \mathcal{L} with scattered missing beats. The robustness of MHR using both \mathcal{L} and \mathcal{NL} gap filling supports that the number of missing beats is well approximated by these methods. Gap-filling degradation with bursts of missing beats is easily explained by the lack of information as the correction moves away from the edges of the burst. Phenomena such as respiratory sinus arrhythmia cannot be inferred in large bursts either. MHR proved to be a very robust metric in missing data scenarios, assuming a worst-case maximum deviation of 0.7 beats per minute. Although not shown in the results section, MHR was able to withstand losses in bursts of up to one minute without the median error exceeding 1 beat per minute. However, it is not easy to establish a threshold for which it is preferable to reject the segment. This rather depends on the stationarity of the data. Because of the metric robustness, in periods where variations are expected, the rate of these changes should be a more dominant factor than metric degradation in the segment rejection decision. The case of scattered losses can be more complex, as depending on the distribution it can be complicated for an outlier detection method to work correctly. This is magnified in cases with large respiratory sinus arrhythmia oscillations. Segment rejection is encouraged when computing RMSSD with $> 25\%$ missing beats, as the third quartile error is around 20%. In any case, attempting to correct segments with more than 35% missing beats or a 20-second burst is not considered adequate.

5.2 Frequency-domain metrics computed via Fourier transform

Regarding frequency metrics calculated via Fourier transform, gap-filling methods show a clear advantage in terms of error and state discrimination. \mathcal{NL} was the best correction method for datasets with scattered missing beats predominance. In datasets with burst predominance, \mathcal{NL} performed better for P_{LF} , while \mathcal{L} obtained better results for P_{HF} . The third quartile of P_{LF} error is greater than 20% in case of segments with more than 25% scattered missing beats, suggesting that those segments should be discarded for P_{LF} analysis. In case of P_{HF} , segments with more than 15% missing beats should also be discarded. Discarding segments is suggested when analyzing missing data in bursts longer than 10 seconds.

Figure 38 shows an example of the \mathcal{M} correction method and \mathcal{L} gap filling of how spectra undergo controlled degradation with increasing loss rates. \mathcal{NL} it is not shown because of the similarity to \mathcal{L} . Interpolation could be the cause of a low-pass effect that can be observed in the spectra as a lower P_{HF} with an increasing loss rate. Gap-filling methods seem to be more robust against this effect. In the case of bursts (Figure 39), the same effect can be observed. Note that the higher degradation of the LF band in this particular example case is not consistent with other examples, as can be seen in the overall degradation results of P_{LF} in Table 6.

5.3 Frequency-domain metrics computed via Lomb-Scargle periodograms

In regards to Lomb-Scargle periodograms, \mathcal{NL} obtained the best results for scattered missing beats. In datasets with burst predominance, \mathcal{L} obtained the best results for P_{HF} , while \mathcal{NL} obtained the best results for P_{LF} . Segment rejection for P_{LF} analysis is suggested with more than 25% scattered missing beats. In the case of P_{HF} analysis, rejection is suggested with more than 15% missing beats.

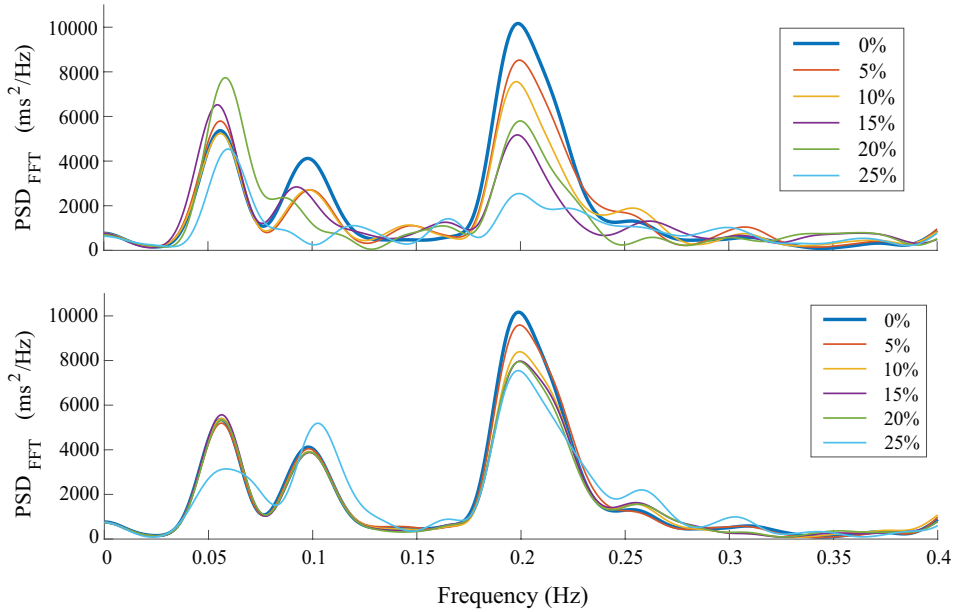


Figure 38: Fourier transform spectral estimates from the same record with different loss rates of scattered missing beats. *Top:* \mathcal{M} correction. *Bottom:* \mathcal{L} gap filling.

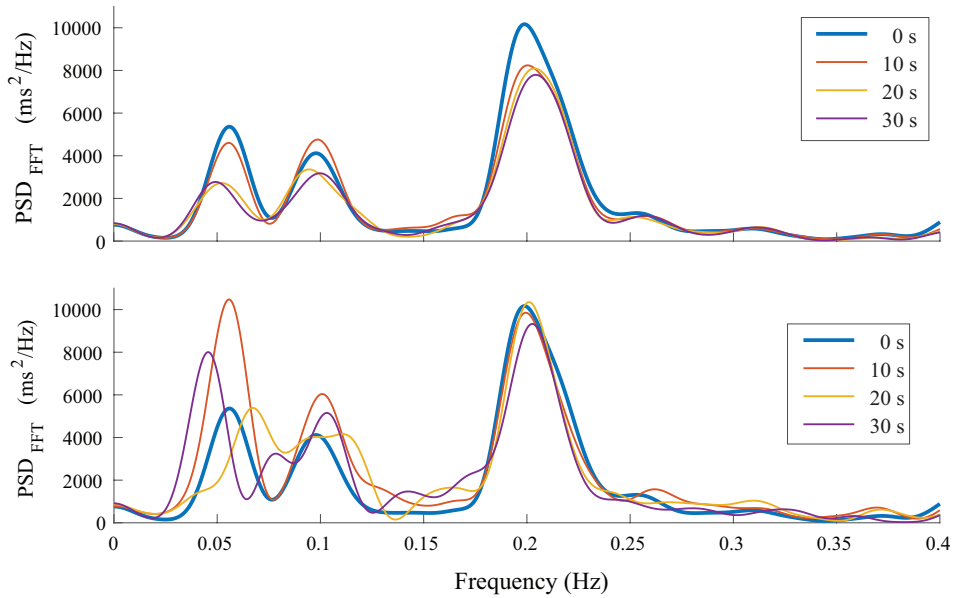


Figure 39: Fourier transform spectral estimates from the same record with different burst durations. *Top:* \mathcal{M} correction. *Bottom:* \mathcal{L} gap filling.

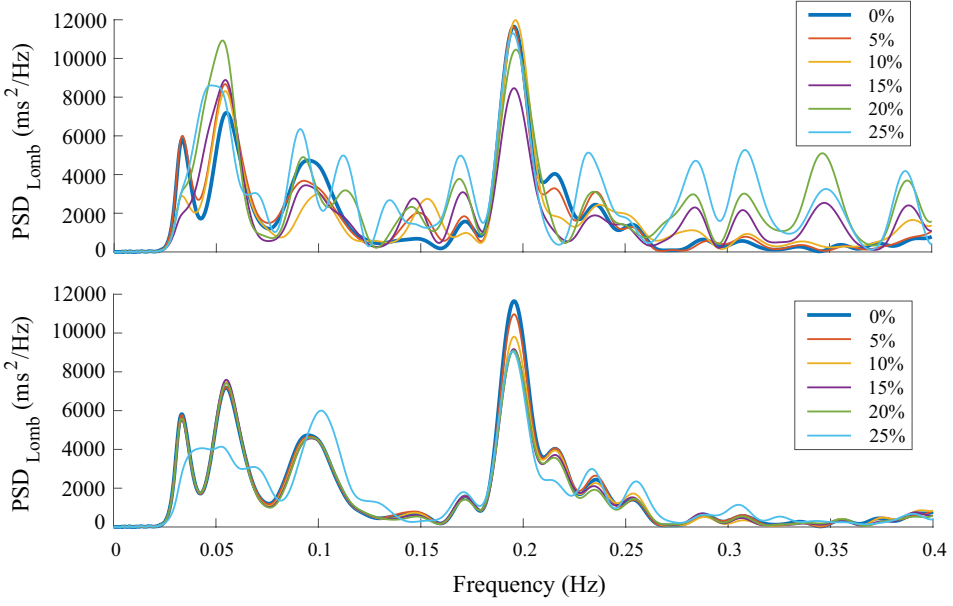


Figure 40: Lomb-Scargle spectral estimates from the same record with different loss rates of scattered missing beats. *Top:* \mathcal{OR} correction. *Bottom:* \mathcal{L} gap filling.

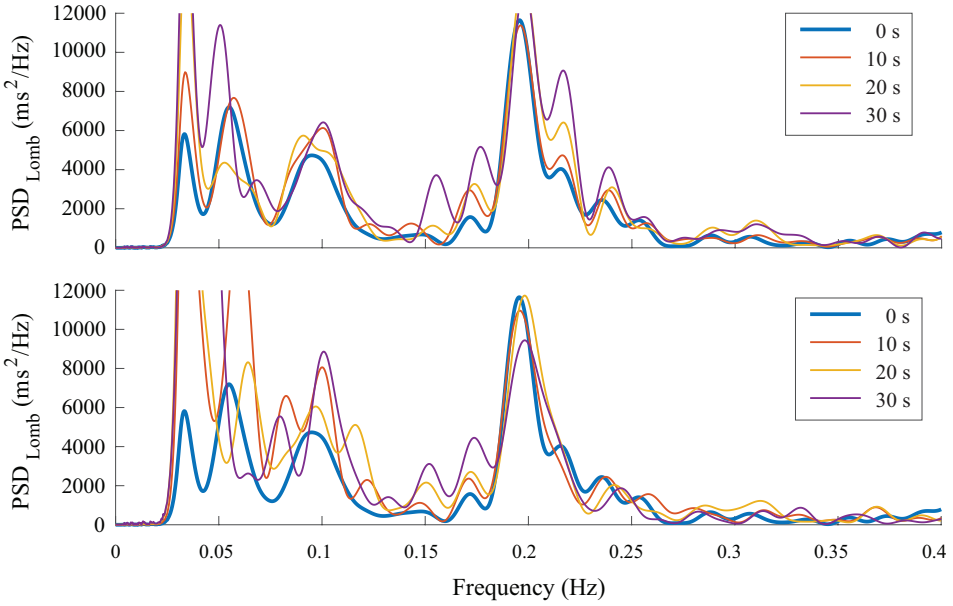


Figure 41: Lomb-Scargle spectral estimates from the same record with different burst durations. *Top:* \mathcal{OR} correction. *Bottom:* \mathcal{L} gap filling.

Segments should be discarded for bursts longer than 10 s as well. Although Lomb-Scargle periodograms allow their use without gap filling—in fact, with no interpolation at all—it deteriorates rapidly in the absence of the whole series (\mathcal{OR} case). This is explained due to the phenomenon of over-oscillation of the spectrum as samples are discarded (see Figures 40 and 41), whose effect is limited when calculating the power by integrating [27], but still causes degradation of the metrics.

5.4 Poincaré Plots

In the case of Poincaré metrics, \mathcal{NL} obtained better results for scattered losses for most metrics, in terms of both error and discrimination between states. \mathcal{L} obtained the best results when analyzing specifically SD1 up to 25% of missing beats, while \mathcal{OR} obtained better results with more than 25% missing beats. However, the third quartile error is greater than 20% in this case, and segment rejection is suggested. As in the case of time-domain metrics, the criterion for rejecting a segment should prioritize the expected stationarity, given the robustness of the metrics with correction methods. \mathcal{OR} obtained the best results in the case of bursts for all metrics.

5.5 Iterative gap-filling approach

The proposed gap-filling method, especially in its non-linear version, has demonstrated to be a very effective correction method. In [15] it was pointed out the difference between correcting the interval series, as it is the case of most of the methods in the literature; and correcting the event series, *i.e.*, the beat-occurrence timestamps. Correcting the interval series involves shifting the timestamps of subsequent beats to address the interval correction. This ultimately means forgoing the reference provided by the subsequent, well-detected beats. Instead, the proposed method corrects the event series without this shifting, by adding a variable number of beats taking into account the budget of seconds to be filled in. Larger gaps require a greater number of filling beats to obtain IBIs in accordance with the

adjacent intervals to the gap.

In [15] it is shown that event correction yields more accurate results than interval series correction. Besides, a novel aspect of the proposed gap-filling method lies in the way in which the correction of each segment is approached. The proposed method is a segment-based iterative algorithm instead of a gap-based one. The use of this kind of algorithm aims to cope with two major problems of event series gap filling: distinguishing outliers at high loss rates and the lack of knowledge of the number of missing beats per gap. Thus, it starts by solving simple gaps before those involving more than one beat. This is an improvement over the majority of gap-filling methods in the literature, where each gap is corrected before moving on to the next one, missing the advantage of solving the shorter gaps first.

It should be noted that the best method is not necessarily the one with the lowest error. Depending on the application, especially working with devices with limited computational capacity and/or that are battery-operated, a method with acceptable results is interesting if it means an improvement in computation time and overall processing load.

5.6 Limitations

Regarding the limitations of this work, it is important to note that this research only focuses on data losses (false negatives in beat detections) and not on general errors (a combination of false positives and false negatives). The presence of false positives has a deleterious impact when trying to obtain the most accurate metrics. This type of error introduces an additional variable: the baseline from which to infer false negatives could be lost. In addition to a previous artifact detection stage, a false beat detection rejection stage should be implemented before applying the presented methods to deal with missing data. If the number of false beat detections is not very high, a moving-average-based algorithm may be enough. This concept is of paramount importance when dealing with wearable devices, especially those that monitor 24/7, since a high percentage of the time, beat detections can be unreliable, and therefore affect any further processing.

Another limitation is the monotonicity of Hermite polynomials. As this interpolation eliminates relative maxima and minima within the burst, it should be taken into account in cases with long bursts and high variability, such as in cases with strong respiratory sinus arrhythmia. Despite this, it performs better than other traditional interpolation methods in the literature such as cubic splines, which present convergence problems by introducing unwanted oscillations. Further work should be done to address this, for example, by introducing estimated stationary points before interpolating. In addition, interpolation methods that do not impose monotonicity while limiting overshooting should also be investigated.

Also, in contrast to the simulation database, respiratory frequencies have not been tested for the Apple Watch database. Therefore, the use of classical frequency bands may result in an incorrect evaluation of the frequency metrics in some cases, and their behavior may differ from that seen in simulation [26]. However, data presentation in medians and quartiles should limit the effect of these outliers on the results.

6 Conclusions

A segment-based gap-filling method for HRV series analysis in the presence of missing data has been presented. Correction is made on the event series, allowing this method to be used independently of the signal used for beat detection (ECG, PPG, etc.). The best-performing correction methodology depends on the analyzed HRV metrics: correction without gap filling is the best option for SDNN, RMSSD, and Poincaré plot metrics in situations when the missing beats are mainly in bursts, whereas they benefit from gap-filling approaches in the cases of scattered missing beats. Gap-filling approaches obtained the best performance in terms of frequency-domain metrics. Furthermore, the performance analysis assists in drawing some conclusions about when to discard a segment for further analysis depending on how much error is assumable in the specific application: in order to obtain estimations with an error lower than 20%, those segments with more than

35% of missing beats or more than 20-second bursts should be discarded for HRV time-domain metrics and Poincaré plots. Moreover, those segments with more than 25% of missing beats or more than 10-second bursts should be discarded for HRV frequency-domain analysis.

Bibliography

- [1] Jerzy Sacha and Władysław Pluta. Alterations of an average heart rate change heart rate variability due to mathematical reasons. *International Journal of Cardiology*, 128(3):444–447, 8 2008.
- [2] Jerzy Sacha. Interplay between heart rate and its variability: A prognostic game. *Frontiers in Physiology*, 5, 2014.
- [3] George E. Billman, Heikki V. Huikuri, Jerzy Sacha, and Karin Trimmel. An introduction to heart rate variability: Methodological considerations and clinical applications. *Frontiers in Physiology*, 6, 2015.
- [4] Martin Buchheit. Monitoring training status with HR measures: Do all roads lead to Rome? *Frontiers in Physiology*, 5, 2014.
- [5] Gaetano Valenza, Ronald G. Garcia, Luca Citi, Enzo P. Scilingo, Carlos A. Tomaz, and Riccardo Barbieri. Nonlinear digital signal processing in mental health: Characterization of major depression using instantaneous entropy measures of heartbeat dynamics. *Frontiers in Physiology*, 6, 2015.
- [6] Matthias Weippert, Kristin Behrens, Annika Rieger, Regina Stoll, and Steffi Kreuzfeld. Heart rate variability and blood pressure during dynamic and static exercise at similar heart rate levels. *PLoS ONE*, 8(12), 12 2013.
- [7] Javier Mateo and Pablo Laguna. Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal. *IEEE Transactions on Biomedical Engineering*, 50(3):334–343, 3 2003.

- [8] J. McNames, T. Thong, and M. Aboy. Impulse rejection filter for artifact removal in spectral analysis of biomedical signals. In *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, volume 26 I, pages 145–148, 2004.
- [9] Kathi J Kemper, Craig Hamilton, and Mike Atkinson. Heart Rate Variability: Impact of Differences in Outlier Identification and Management Strategies on Common Measures in Three Clinical Populations. *Pediatric research*, 62(3):337–342, 2007.
- [10] Ming Yuan Lee and Sung Nien Yu. Improving discriminability in Heart Rate variability analysis using simple artifact and trend removal preprocessors. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10*, pages 4574–4577, 2010.
- [11] S. Begum, M. S. Islam, M. U. Ahmed, and P. Funk. K-NN based interpolation to handle artifacts for heart rate variability analysis. In *IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2011*, pages 387–392, 2011.
- [12] Hussein Al Osman, Mohamad Eid, and Abdulmotalieb El Saddik. A Pattern-Based Windowed Impulse Rejection Filter for Nonpathological HRV Artifacts Correction. *IEEE Transactions on Instrumentation and Measurement*, 64(7):1944–1957, 7 2015.
- [13] David A Giles and Nick Draper. Heart Rate Variability During Exercise: A Comparison of Artefact Correction Methods. *Journal of Strength and Conditioning Research*, 32(3):726–735, 2017.
- [14] Hyun Jae Baek and Jae Wook Shin. Effect of Missing Inter-Beat Interval Data on Heart Rate Variability Analysis Using Wrist-Worn Wearables. *Journal of Medical Systems*, 41(10), 10 2017.
- [15] Davide Morelli, Alessio Rossi, Massimo Cairo, and David A. Clifton. Analysis of the impact of interpolation methods of missing RR-intervals caused by motion artifacts on HRV features estimations. *Sensors*, 19(14), 7 2019.

- [16] Mouna Benchekroun, Baptiste Chevallier, Dan Istrate, Vincent Zalc, and Dominique Lenne. Preprocessing Methods for Ambulatory HRV Analysis Based on HRV Distribution, Variability and Characteristics (DVC). *Sensors*, 22(5), 3 2022.
- [17] Aleksandra Królak, Tomasz Wiktorski, Magnus Friestad Bjørkavoll-Bergseth, and Stein Ørn. Artifact correction in short-term hrv during strenuous physical exercise. *Sensors (Switzerland)*, 20(21):1–23, 11 2020.
- [18] Deepak Berwal, C. R. Vandana, Sourya Dewan, C. V. Jiji, and Maryam Shojaei Baghini. Motion Artifact Removal in Ambulatory ECG Signal for Heart Rate Variability Analysis. *IEEE Sensors Journal*, 19(24):12432–12442, 12 2019.
- [19] Ayca Aygun, Hassan Ghasemzadeh, and Roozbeh Jafari. Robust interbeat interval and heart rate variability estimation method from various morphological features using wearable sensors. *IEEE Journal of Biomedical and Health Informatics*, 24(8):2238–2250, 8 2020.
- [20] E. Gil, M. Orini, R. Bailón, J. M. Vergara, L. Mainardi, and P. Laguna. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. *Physiological Measurement*, 31(9):1271–1290, 2010.
- [21] M. Pagani, F. Lombardi, S. Guzzetti, O. Rimoldi, R. Furlan, P. Pizzinelli, G. Sandrone, G. Malfatto, S. Dell’Orto, and E. Piccaluga. Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circulation Research*, 59(2):178–193, 1986.
- [22] N. Montano, T.G. Ruscone, A. Porta, F. Lombardi, M. Pagani, and A. Malliani. Power spectrum analysis of heart rate variability to assess the changes in sympathovagal balance during graded orthostatic tilt. *Circulation*, 90(4 I):1826–1831, 1994.
- [23] Juan Pablo Martínez, Rute Almeida, Salvador Olmos, Ana Paula Rocha, and Pablo Laguna. A Wavelet-Based ECG Delineator Evaluation

- on Standard Databases. *IEEE Transactions on Biomedical Engineering*, 51(4):570–581, 4 2004.
- [24] David Hernando, Surya Roca, Jorge Sancho, Álvaro Alesanco, and Raquel Bailón. Validation of the apple watch for heart rate variability measurements during relax and mental stress in healthy subjects. *Sensors*, 18(8), 8 2018.
- [25] Ko Keun Kim, Jung Soo Kim, Yong Gyu Lim, and Kwang Suk Park. The effect of missing RR-interval data on heart rate variability analysis in the frequency domain. *Physiological Measurement*, 30(10):1039–1050, 10 2009.
- [26] Alberto Hernando, Jesús Lázaro, Eduardo Gil, Adriana Arza, Jorge Mario Garzón, Raúl López-Antón, Concepción De La Camara, Pablo Laguna, Jordi Aguiló, and Raquel Bailón. Inclusion of Respiratory Frequency Information in Heart Rate Variability Analysis for Stress Assessment. *IEEE Journal of Biomedical and Health Informatics*, 20(4):1016–1025, 7 2016.
- [27] Pablo Laguna, George B Moody, Associate Member, Roger G Mark, and Senior Member. Power Spectral Density of Unevenly Sampled Data by Least-Square Analysis: Performance and Application to Heart Rate Signals. *IEEE Transactions on Biomedical Engineering*, 45(6):698–715, 1998.

Mental health monitoring using wearables

1 Objectives and motivations

This study evaluates the use of smartphone-camera-acquired photoplethysmography (SCPPG) to derive PRV metrics for monitoring depression and anxiety. SCPPG signals may exhibit lower quality compared to those obtained from conventional PPG devices. Movements, along with variations in finger pressure on the camera, are the primary factors causing artifacts. Therefore, validation for different applications must be conducted within each specific scenario. SCPPG has previously been validated for obtaining PRV metrics [1–3] and has been utilized in wellness and healthcare applications [4–7]. However, to the best of the authors' knowledge, this study represents the first instance of SCPPG being validated in a protocol specifically designed to monitor anxiety and depression.

This study includes a comparison between SCPPG and a validated device within a stress assessment protocol, administered to individuals diagnosed with MDD and/or GAD, as well as a control group. The aim is to determine whether the ANS-related metrics obtained from the SCPPG are consistent with those obtained from the reference device. An initial study demonstrated a strong concordance of PRV metrics in a limited group of healthy individuals [8]. Additionally, this research aims to further evaluate whether SCPPG can accurately identify stress reactivity and if the variations between

the patient and control groups align with those observed using the reference device. Successful validation would add a broadly accessible tool to the existing methods for monitoring anxiety and depression.

The novelty of this work is the validation of the PRV metrics using SCPPG in a protocol designed for monitoring anxiety and depression, including tasks involving movement and speech (see Section 2.1). These metrics include the sympathovagal balance obtained by orthogonal subspace decomposition of the heart rate modulating signal from the SCPPG and a respiration signal (see Section 3.3), which enhances the accuracy of HRV analysis by separating respiration-related fluctuations [9].

2 Experiment

A total of 82 individuals participated in a depression assessment protocol. Among them, 24 were diagnosed with GAD and/or MDD (patient group), whereas 58 were healthy individuals (control group). Exclusion criteria ensured that participants in both groups did not have any cardiac, neurological, or endocrine conditions, nor any other psychiatric disorders, to prevent potential confounders. Participants taking tricyclic antidepressants, beta-blockers, and antipsychotics were also excluded due to their effects on the autonomic control of the heart [10]. Out of the 82 participants, 3 individuals (2 patients and 1 control) experienced difficulties with finger stability on the camera due to tremors. This issue is addressed in Section 5. As a result, these participants were excluded from subsequent analyses, leaving a final sample size of 79 subjects, consisting of 22 patients and 57 controls. Patients were recruited by the Hospital Clínico Universitario (Zaragoza, Spain) and by SOS Adolescentes (Huesca, Spain), while controls were recruited by social networks.

During the procedure, participants were directed to remain seated and reduce their movements. Utilizing their non-dominant hand, they held a Xiaomi Pocophone F1 smartphone (Xiaomi Inc., China) and placed their index finger over the camera. This smartphone's camera positioning let

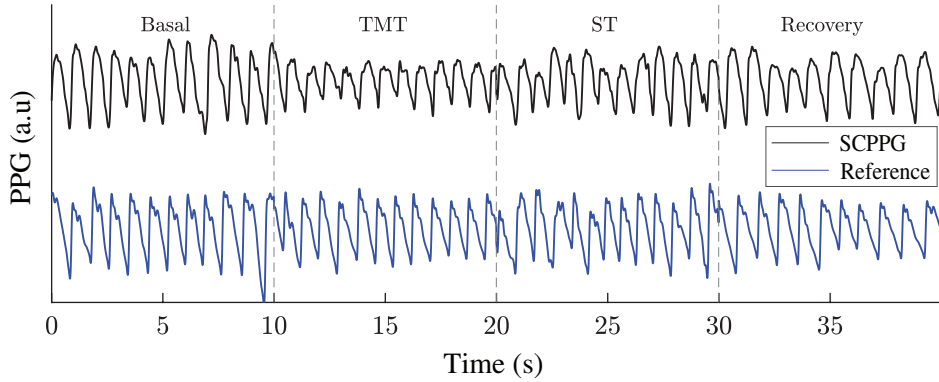


Figure 42: SCPPG (black) and Reference (blue) signals. The image consists of 10-second segments extracted from different stages of the protocol, which have been assembled together. The time axis is shown as a reference; it does not indicate continuity.

users grip the device comfortably without touching the flashlight directly. The flashlight is positioned about 5 mm to the right of the camera, which effectively averts any discomfort due to heat during extended periods of recording. The flashlight remained on during the recordings, and the autofocus and autoexposure functions of the camera were disabled to prevent non-physiological oscillations in the SCPPG signal. This step is essential since such oscillations usually occur at a frequency similar to that of the blood pulses and potentially cause confusion.

A Medicom system (Medicom MTD, Russia), hereinafter referred to as the Reference, was simultaneously used to record a conventional PPG signal from the ring finger of the same hand for comparison, with a sampling rate of 250 Hz (see Figure 42). Besides this PPG signal, this device also recorded the respiratory effort from the chest at the same sampling rate. The smartphone app used was created in-house to record SCPPG data. The software analyzes the video stream, which has a resolution of 320x240 pixels and uses RGB encoding, by summing the green component values of each frame to produce a signal directly related to the image's green intensity. Using the green channel yields a more robust cardiac pulse signal compared to the red or blue bands due to hemoglobin's significant absorption in the green spectrum [1, 11, 12].

The frame rate is approximately 24 frames per second, although it can vary based on the operating system's workload. The signal is then upsampled to 250 Hz using cubic splines to match the reference sample rate. Interpolation and further processing were conducted offline using MATLAB (MathWorks, USA).

Before starting the stress protocol, the participants were interviewed to gather their personal information and confirm the inclusion and exclusion criteria. During this interview, participants completed four tests to evaluate their self-perceived stress: the Perceived Stress Scale (PSS), the State-Trait Anxiety Inventory (STAI), which is divided into state (S-STAI) and trait (T-STAI) components, and two custom tests named ES3 and ES3-V [13].

The population data and test results are shown in Table 16. All 79 participants were included in the validation of SCPPG's agreement with the reference. Following this, SCPPG's capability to assess stress reactivity was evaluated. For this evaluation, reactivity was compared to the reference, and the control and patient groups were separated. Subjects from the control group were matched with those from the patient group based on age, sex, and body mass index (BMI). A total of 22 pairs were matched: 5 pairs of men and 17 pairs of women. The inclusion of these data sets was approved by the ethical committee of the Gobierno de Aragón (CEICA, PI20/430), and all patients signed an informed consent.

2.1 Stress protocol

The protocol encompassed the following stages: i) a 5-minute Basal state; ii) the Trail Making Test (TMT); iii) the Stroop Test (ST); and iv) a 5-minute Recovery phase. During the Basal stage, participants were instructed to relax with the aid of an audio guide. Subsequently, they engaged in the TMT, beginning with a page featuring a series of numbers scattered randomly, which participants were required to connect in ascending order using the index finger of their dominant hand, ensuring continuous contact with the tablet. The subsequent page presented a combination of numbers and letters, alternating between the two elements in a sequence (1-A-2-B-3-C,

and so forth).

The ST comprised three pages. On the first page, participants encountered the words "red," "green," and "blue," which were presented in black ink in a random order for them to read aloud. The second page featured colored inks —red, green, and blue— that participants were required to identify and name. On the third page, participants were presented with the same set of words but written in ink colors that did not match the word itself, requiring them to state the color of the ink rather than the word. The Recovery phase involved unguided relaxation immediately after the ST.

Table 16: Demographic characteristics. Significant differences between paired controls and patients ($p < 0.05$) in bold. *SD* = Standard Deviation; *N* = Number; BMI = Body Mass Index; PSS = Perceived Stress Scale; STAI = State-Trait Anxiety Inventory.

	Group		
	All subjects (<i>n</i> = 79)	Paired controls (<i>n</i> = 22)	Patient (<i>n</i> = 22)
Age, Mean (<i>SD</i>)	31 (12)	30 (10)	30 (12)
Age, Range	18-66	19-64	18-64
Male, <i>N</i> (%)	37 (47%)	5 (23%)	5 (23%)
Weight [kg], Mean (<i>SD</i>)	70 (13)	65 (10)	66 (12)
Height [cm], Mean (<i>SD</i>)	172 (9)	168 (8)	169 (10)
BMI, Mean (<i>SD</i>)	24 (3)	23 (3)	23 (4)
PSS, Mean (<i>SD</i>)	15.2 (6.4)	14.3 (5.8)	20.0 (7.5)
PSS, Range	5-32	5-20	5-32
S-STAI, Mean (<i>SD</i>)	17.7 (9.8)	14.9 (6.3)	24.3 (12.2)
S-STAI, Range	2-50	2-21	6-50
T-STAI, Mean (<i>SD</i>)	19.7 (10.7)	16.0 (7.3)	28.7 (12.7)
T-STAI, Range	3-52	3-24	9-52
ES3, Mean (<i>SD</i>)	17.7 (17.0)	8.8 (6.1)	31.5 (21.0)
ES3, Range	0-69	0-23	2-69
ES3-V, Mean (<i>SD</i>)	37.6 (24.7)	37.4 (23.9)	45.6 (28.4)
ES3-V, Range	0-95	3-65	0-95

Note: PSS ranges from 0 to 40, S-STAI and T-STAI range from 0 to 60, ES3 ranges from 0 to 80, and ES3-V ranges from 0 to 100.

3 Methods

3.1 Filtering and artifact removal

The Reference PPG and the SCPPG are subjected to third-order Butterworth high-pass filtering at 0.3 Hz and low-pass filtering at 10 Hz. In order to reduce motion artifacts, which often affect SCPPG signals and lead to incorrect or missed pulse detections, an artifact detector is used. This detector sets upper and lower limits for the Hjorth parameters (see Chapter 4, Section 2.1). Hjorth parameters are calculated from signal data, either the Reference or the SCPPG, within 4-second sliding windows ($P = 1000$ samples) with a 25% overlap.

3.2 Pulse detection

Pulses are identified using the adaptive threshold algorithm described in Chapter 4, Section 4.2. This algorithm provides the event series (t_k) , which represents the timestamps of pulse occurrences at their maximum upslope. Before performing a PRV analysis, it is essential to identify and correct any misdetection.

To achieve this, the interval function $d_{IF}(t)$ is computed (see Equation 4.10), and a moving median of 30 samples is used to detect outliers. The moving median produces an expected pulse-to-pulse interval d_{EPPI} at each t_k , calculated analogously to $d_{EIBI}(t_k)$ of Equation 4.11. The interval at t_k is considered a false positive if $d_{IF}(t_k) < (0.7 \times d_{EPPI}(t_k))$. These false positives are removed from the t_k series, and $d_{IF}(t_k)$ is computed again. Next, the intervals at t_k are considered false negatives if $d_{IF}(t_k) > (1.3 \times d_{EPPI}(t_k))$. The treatment of false negatives is not the same for all PRV metrics, following the recommendations of Chapter 6. This treatment is detailed in Section 3.3.

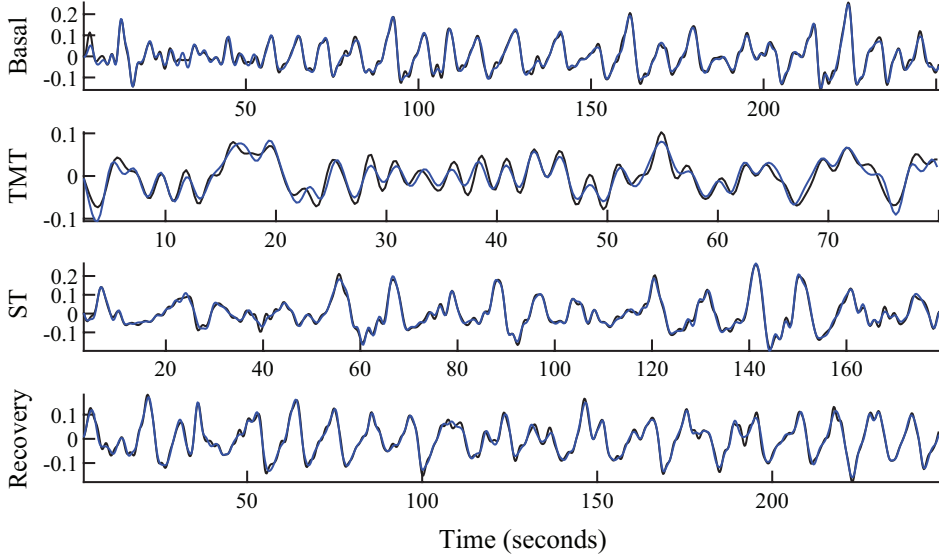


Figure 43: SCPPG (black) and Reference (blue) modulating signals $m(t)$ obtained with the TVIPFM model for each stage. The agreement is very high at all stages, especially at rest.

3.3 PRV metrics

Three time-domain and three frequency-domain metrics were calculated. The metrics in the time domain were the MHR, the SDNN, and the RMSSD. Gap filling of false negatives is not recommended for these metrics. Therefore, they were calculated from $d_{\text{IF}}(t_k)$, excluding outliers in the t_k series.

Conversely, prior to calculating metrics in the frequency domain, the gaps were addressed using the \mathcal{NL} gap-filling method. In addition, an exploratory analysis revealed that respiratory rate often appeared within the LF range, *i.e.*, 0.04 to 0.14 Hz, prompting an improved analysis of HRV metrics based on OSP (see Chapter 4, Section 6.3). The heart rate modulating signal, $m(t)$, is calculated at 4 Hz using the TVIPFM model (see Figure 43). Then, $m(t)$ is decomposed by the respiratory signal, $r(t)$, into a respiration-related linear component, $\hat{m}_r(t)$, and a residual component, $\hat{m}_\perp(t)$, containing other modulators. Subsequently, $\hat{m}_r(t)$ and $\hat{m}_\perp(t)$ are detrended with a 4th order Butterworth high-pass filter with a cutoff frequency 0.03 Hz. Spectral densities $\hat{S}_{m_r}(f)$ and $\hat{S}_{m_\perp}(f)$ are estimated using Welch periodograms with

120-second Hamming windows and 30-second overlap within all available data in each phase of the protocol. Sympathovagal balance \mathcal{R}' was computed as described in Equation 4.33.

The metrics are evaluated for physiological verisimilitude. Permissible ranges for these metrics are set at $[40,180]$ bpm for MHR; $[5,140]$ ms for SDNN and RMSSD; $[0,0.003]$ *a.u.* for $P_{\perp,LF}$; and $[0,0.05]$ *a.u.* for $P_{r,LF+HF}$. These ranges were derived from the highest and lowest values observed with the reference, including an ample allowance. Any metrics that fall outside the specified ranges are identified and removed from the agreement and reactivity analysis. Instead, the count of these cases is presented.

3.4 Statistical analysis

PRV metrics derived from the SCPPG are compared to those from the Reference. The Pearson correlation (r) is evaluated with a significance threshold of $\alpha = 0.05$. Bland-Altman's mean difference (Δ), given as a percentage with its 95% confidence interval, is used to assess bias. To determine if the reactivity of controls and patients comes from probability distributions with different medians, a Wilcoxon rank-sum test is conducted assuming nonnormal distributions. This assumption is supported by the presence of both normal and non-normal distributions across different stages, as determined by a Shapiro-Wilk test. Although subjects in the control and patient groups are matched, as detailed in Section 2, the statistical test is unpaired. The effect size of the reactivity is measured using Cliff's Delta [14].

4 Results

4.1 Agreement

Table 17 presents the percentages of artifact-free time as determined by the Hjorth-based automatic detector. The stability of the SCPPG is demonstrated with 98.8% of the time being artifact-free, which is only 0.7% lower than

the Reference value. Across different stages, no significant differences are observed, with a decrease of only 2.4% between the Basal and ST stages, representing the maximum and minimum values, respectively. The minimum value in the Reference is also observed in the ST stage, though it is only 1.2% lower than the Basal value.

The PRV values, along with the corresponding absolute errors, Pearson correlation coefficients, and the bias relative to the Reference, are presented in Table 18. The last column represents the number of missing values. The errors are at least an order of magnitude smaller than the values of the metrics. Notably, the MHR error is exceptionally small, with an error magnitude three orders smaller than the metric value. All metrics exhibit high correlation, with coefficients exceeding 0.98, except for the RMSSD at the TMT stage, where the correlation is still substantial at 0.89. The p -values for all correlation tests were less than 0.001.

The number of missing values results from both the restrictions in the allowable range (see Section 3.3), indicating a failure in the calculation, and the requirement of a minimum duration of 60 seconds for spectrum computation in frequency-domain metrics. For the time-domain metrics, the instances of unreliable measurements are less than 2.5%. In contrast, frequency-domain metrics display higher percentages of missing values, with the TMT stage reaching up to 22.8%. Excluding the TMT stage, the percentage of missing cases is less than 6.4%.

The bias is nearly negligible across all metrics except for RMSSD. This

Table 17: Percentage of artifact-free time according to Hjorth-based automatic detection.

Stage	SCPPG (%)	Reference (%)
Basal	99.5	99.9
TMT	99.4	99.3
ST	97.1	98.7
Recovery	99.0	99.8
Total	98.8	99.5

particular metric shows a steady positive bias in all phases, ranging from 13.6% to 21.2%. Figure 44 presents the linear regressions and Bland-Altman plots for all combined stages. Every correlation has a p -value below 0.001, with the lowest correlation being 0.96. Once more, only the RMSSD exhibits a significant bias, though the overall effect of the stages diminishes this bias to 12%.

Table 18: Agreement results of PRV metrics. Median and standard deviation, error median and standard deviation, Pearson correlation coefficient, number of missing cases, and Bland-Altman mean difference. All subjects.

	Values, Median (SD)	Error, Median (SD)	Pearson's r	Bland-Altman's Δ , M [C.I.] (%)	Missing, N (%)
MHR [bpm]					
Basal	72 (11)	0.0 (0.1)	1.00	0.0 [-0.2 0.3]	0 (0%)
TMT	77 (13)	0.0 (0.1)	1.00	0.0 [-0.3 0.3]	0 (0%)
ST	80 (15)	0.1 (0.2)	1.00	0.0 [-0.6 0.6]	1 (1.3%)
Recovery	71 (10)	0.0 (0.0)	1.00	0.0 [-0.1 0.2]	0 (0%)
SDNN [ms]					
Basal	66 (26)	1.0 (1.0)	1.00	2.1 [-4.5 8.8]	0 (0%)
TMT	50 (17)	1.5 (1.5)	0.99	3.4 [-6.6 13.4]	1 (1.3%)
ST	57 (19)	1.7 (2.6)	0.99	4.3 [-10.5 19.0]	1 (1.3%)
Recovery	61 (26)	1.0 (1.0)	1.00	2.3 [-3.2 7.8]	0 (0%)
RMSSD [ms]					
Basal	46 (20)	3.8 (3.2)	0.99	13.6 [-18.3 45.4]	2 (2.5%)
TMT	38 (13)	5.5 (5.3)	0.89	19.9 [-18.7 58.4]	0 (0%)
ST	41 (14)	6.1 (6.1)	0.90	21.2 [-16.4 58.8]	1 (1.3%)
Recovery	43 (21)	4.4 (3.6)	0.99	15.1 [-18.0 48.1]	1 (1.3%)
$P_{\perp, \text{LF}}$ [$a.u.$]					
Basal	$42 (46) \times 10^{-5}$	$1 (8) \times 10^{-5}$	0.98	1.4 [-17.7 20.4]	2 (2.5%)
TMT	$29 (34) \times 10^{-5}$	$1 (2) \times 10^{-5}$	1.00	3.8 [-21.9 29.6]	18 (22.8%)
ST	$52 (44) \times 10^{-5}$	$2 (13) \times 10^{-5}$	0.95	3.1 [-40.6 46.7]	5 (6.4%)
Recovery	$47 (49) \times 10^{-5}$	$2 (10) \times 10^{-5}$	0.98	2.1 [-22.1 26.4]	1 (1.3%)
$P_{r, \text{LF} + \text{HF}}$ [$a.u.$]					
Basal	$164 (238) \times 10^{-5}$	$4 (19) \times 10^{-5}$	1.00	0.7 [-14.8 16.3]	0 (0%)
TMT	$31 (48) \times 10^{-5}$	$2 (3) \times 10^{-5}$	1.00	7.5 [-34.9 49.8]	18 (22.8%)
ST	$84 (79) \times 10^{-5}$	$3 (14) \times 10^{-5}$	0.98	0.0 [-27.3 27.3]	5 (6.4%)
Recovery	$92 (228) \times 10^{-5}$	$4 (15) \times 10^{-5}$	1.00	-0.3 [-21.8 21.2]	0 (0%)

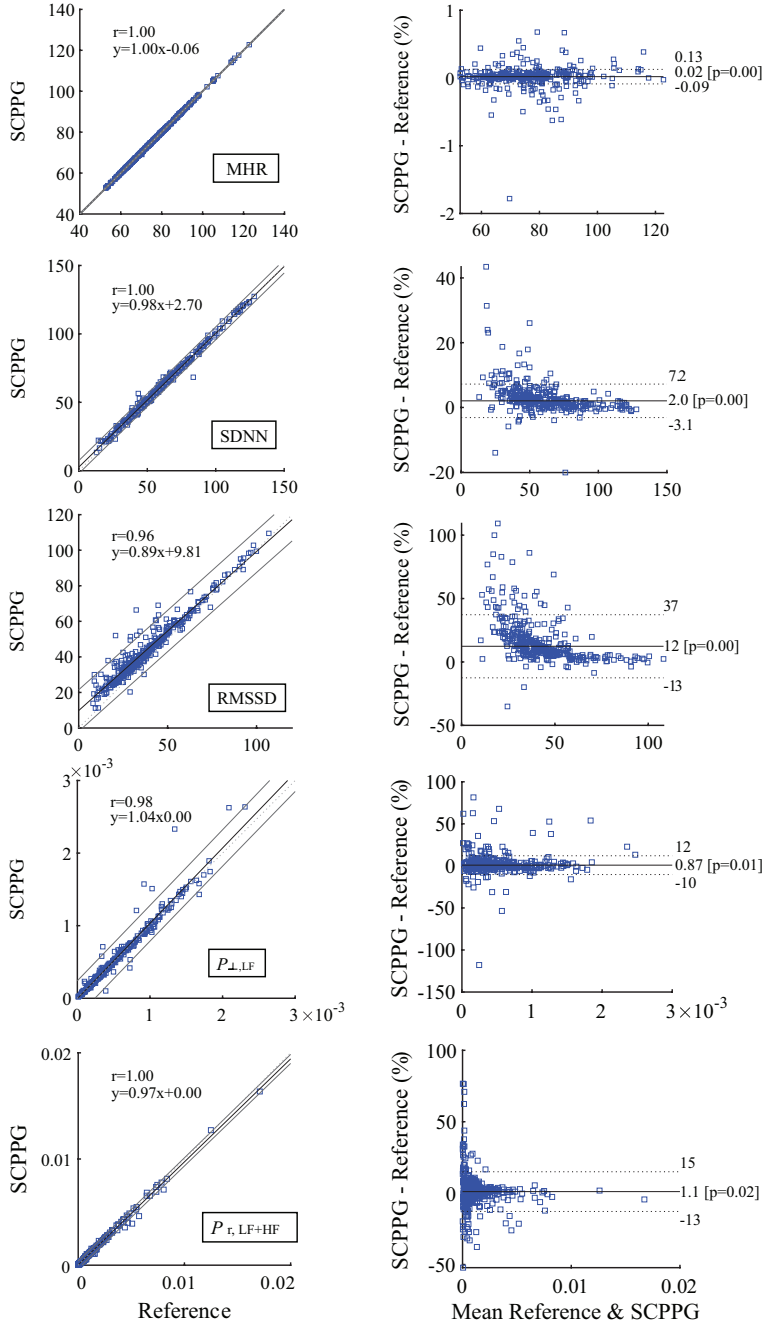
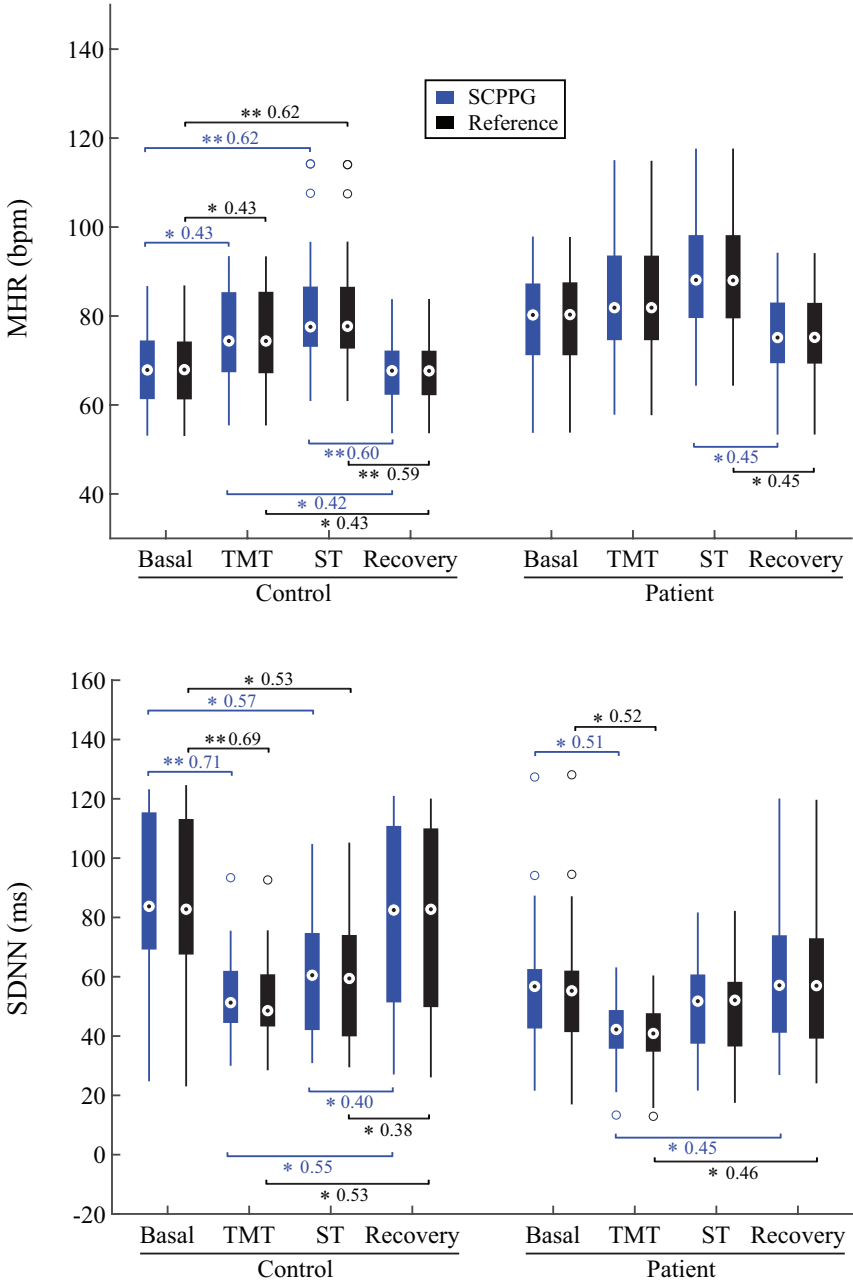


Figure 44: Regression and Bland-Altman plots of relative differences for each metric. All stages together. First: MHR. Second: SDNN. Third: RMSSD. Fourth: $P_{\perp, LF}$. Fifth: $P_{r, LF+HF}$



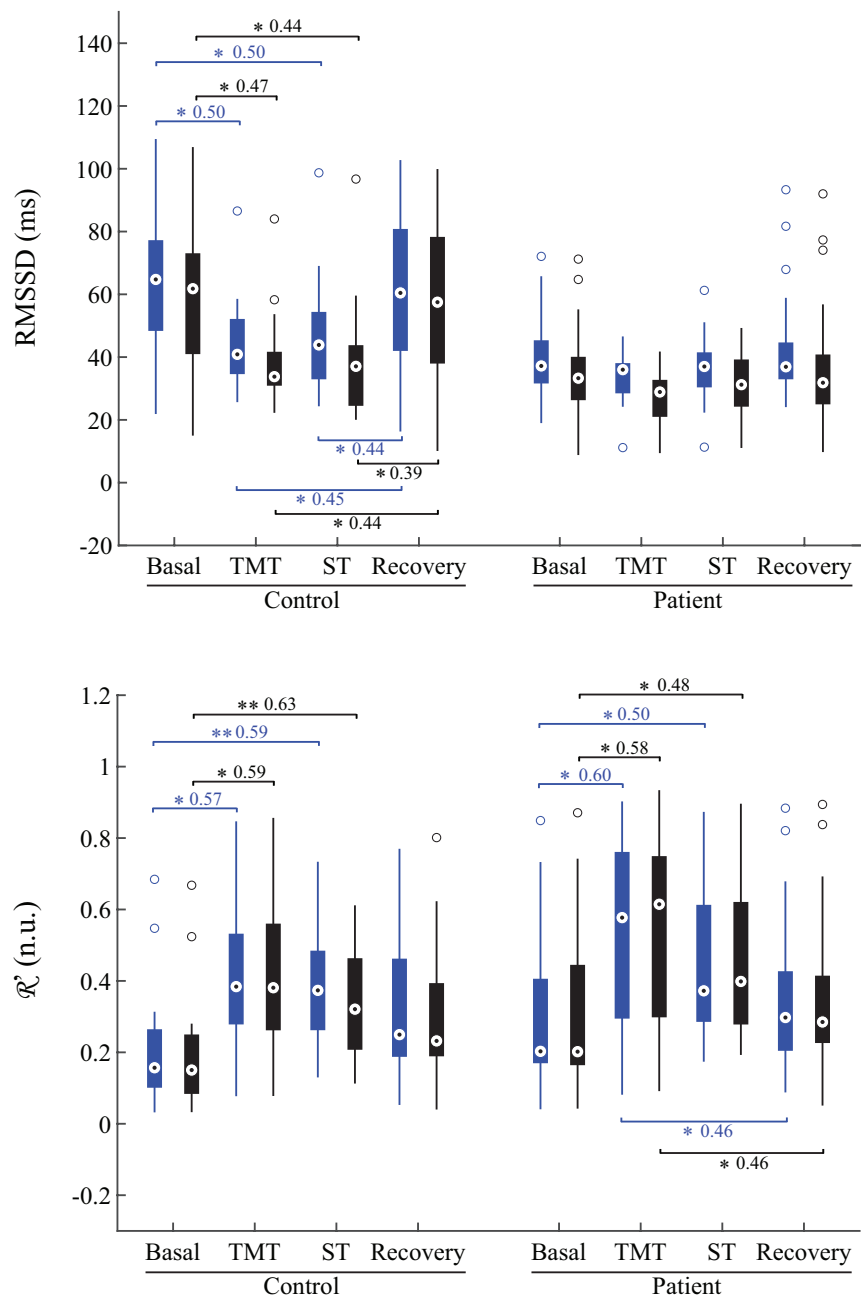


Figure 44: Boxplots for each metric. Significant differences are marked with asterisks: * for $p < 0.05$ and ** for $p < 0.001$. Values after asterisks indicate the effect size. First: MHR. Second: SDNN. Third: RMSSD. Fourth: \mathcal{R}' .

4.2 Stress reactivity

Figure 44 illustrates the distributions of PRV metrics across different stages for both control and patient groups. In the control group, increases in MHR and \mathcal{R}' are observed during stress phases (TMT and ST) compared to relaxation phases (Basal and Recovery). Conversely, SDNN and RMSSD exhibit a decreasing trend. Significant differences and effect sizes are indicated in the graphs. Notably, significant differences were found in all metrics between the Basal stage and the two stress tests in the control group. Additionally, significant differences are present in MHR, SDNN, and RMSSD between the stress tests and the Recovery phase.

In the patient group, reduced stress reactivity is observed. No significant differences are found between the Basal stage and the stress tests in MHR or RMSSD. However, differences in SDNN between Basal and TMT are observed, although with a reduced effect size (from 0.71 to 0.51). Significant differences are also noted in \mathcal{R}' between Basal and both stress tests, with a smaller effect size between Basal and ST. The differences in MHR, RMSSD, and SDNN between the stress tests and the Recovery phase are either not significant or have reduced effect sizes. An exception is found in \mathcal{R}' , where differences are more pronounced in the patient group due to a significant increase during the TMT.

There is a high degree of agreement between the distributions obtained from the SCPPG and the Reference. The statistical tests for both the SCPPG and the Reference show consistent results in terms of statistical significance and effect size when there is good agreement.

The control group completed the TMT in 106 ± 30 seconds, while the patient group took 118 ± 32 seconds. For the ST, the control group finished in 211 ± 31 seconds, compared to 233 ± 57 seconds for the patient group. The differences in median test durations between the groups are not statistically significant ($p = 0.25$ and $p = 0.33$).

5 Discussion

5.1 SCPPG for PRV monitoring

The concordance between the PRV metrics derived from the SCPPG and the reference is exceptionally high. Notably, a correlation exceeding 0.96 ($p < 0.001$) indicates that the smartphone is as reliable as a commercial pulse oximeter, with correctable bias. The capability of the metrics obtained from the SCPPG to identify stress reactivity matches that of the reference device in terms of both statistical significance and effect size. Furthermore, the observed differences in reactivity between the control and patient groups support the use of the SCPPG for monitoring stress reactivity in the context of anxiety and depression assessments. The difference between stress stages and the Recovery was lower in subjects from the patient group, which is in agreement with the results found in [15]. The study of impaired recovery is highly intriguing, as research has connected it to a heightened risk factor for cardiovascular disease [15, 16]. Differential resting levels are evident between the patient and control groups. This is in agreement with the literature and may be a valuable measure, especially in longitudinal studies, given its relationship to the regulation of emotions and attention in addition to physiological regulation [17].

The overestimation of RMSSD may be attributed to the low sampling rate of SCPPG as lower sampling rates are more susceptible to noise, leading to signal distortion and reduced temporal resolution. Choi *et al.* [18] note that RMSSD requires a higher sampling frequency for robustness, unlike MHR, which averages out variability and remains robust. They recommend a minimum sampling frequency of 25 Hz. Similarly, Beres *et al.* [19] suggest that RMSSD demands higher sampling rates compared to metrics like MHR and SDNN, reducing the necessary rate to 20 Hz. Thus, the camera's 24 Hz sampling rate is marginal or potentially insufficient, especially given Android's inconsistent handling of the camera, leading to lower sampling rates. Therefore, future research will explore higher sampling frequencies, supported by newer devices, and non-Android platforms for more stable

sampling rates [20].

Frequency domain metrics are derived through the integration of power spectral density estimates, which are calculated using Welch's method. The durations of the TMT and ST stages are notably shorter compared to those of the Basal and Recovery stages. Specifically, the TMT stage lasted an average of 109 seconds, with a range from 61 to 186 seconds. The ST stage had an average duration of 223 seconds, ranging from 153 to 368 seconds. Consequently, the number of segments available for Welch averaging is reduced in the shorter stages, resulting in increased variance.

Additionally, the brief duration of the TMT stage often leads to a high number of missing cases in frequency domain metrics because it frequently does not meet the minimum requirement of 60 seconds necessary for spectrum calculation. It is important to highlight that although the test may exceed the minimum required duration, artifacts and misdetections can reduce the length of the valid signal used for metric computation, as gaps exceeding 10 seconds are prohibited. When larger gaps are present, the algorithm retains the longest segment without 10-second gaps [21].

The significant increase in \mathcal{R}' observed during the TMT in the patient group might be attributed to the reduction in the number of cases. Although the metrics in the frequency domain are comparable to those in the time domain in terms of both correlation and error, a frequency domain metric could not be obtained in nearly a quarter of the TMT stages. It is noteworthy that only in one instance, the unique missing case in the Recovery stage, the Reference was able to obtain the metric where the SCPPG could not. In all other cases where the SCPPG failed, the Reference also failed.

When using frequency domain metrics, it must be considered that they have more restrictive requirements and that the duration of the test is a crucial factor. Significant differences in HF and LF/HF ratio, calculated without OSP, are reported in the literature [10, 22]. To our knowledge, this study represents the first application of \mathcal{R}' in the surveillance of stress reactivity, overcoming limitations in the respiratory rate.

One of the primary challenges in SCPPG is the significant exposure to the

relative movement of the finger in relation to the camera. This problem is particularly detrimental to patients with tremors, even if the tremors are minor. Three subjects had tremors incompatible with the SCPPG recording. This represents 3.7% of the studied population. The sample is small to extrapolate this percentage to the general population, but it represents an important problem to be solved in the automation of monitoring. Using clamp-type sensors, similar to those used in clinical settings, connected to the smartphone would lead to the loss of the main advantage of this system, which is the widespread availability of sensors in people's pockets. During the recording sessions, participants had to hold the smartphone in the same hand that had a pulse oximeter on the ring finger, leading to an uncomfortable and awkward grip. It is anticipated that the signal quality would be better if the free hand were used.

The percentage of time and cases in which metrics can be obtained is noteworthy. Excluding subjects who were removed due to difficulties in remaining still, the proportion of artifact-free time is 98.78%, which is comparable to the Reference. This high percentage is particularly remarkable given that hand movements are visibly transmitted to the signal. It is likely that this outcome is influenced by the protocol design, which facilitates maintaining a stable and comfortable position. The relaxation stages are performed in complete stillness, and the ST is a spoken task. The only stage that demands movement is the TMT. This test was conducted with the dominant hand while holding the smartphone with the non-dominant hand. This independence contributed to the stability of the recording.

When designing other stress-inducing tasks, particularly those that involve interacting with the recording device—centralizing all interactions on a single device is an ideal scenario—it is recommended to consider that such interaction might introduce artifacts. This matter is currently under investigation.

The stage with the highest number of artifacts was the ST. This was also the stage where subjects reported the highest stress. This, combined with the longer duration of the test and the consequent fatigue, suggests that the induced stress level is more determinant for the quality of the signals

than the manner of interacting with the device. An interesting line of work involves the interplay between human-device interaction design and stress-inducing tasks that facilitate functionality.

It should be noted that artifact-free time has been measured with an automatic detector. This does not imply that the non-rejected segments have a good signal-to-noise ratio nor that the detections are error-free. The Hjorth-based detector rejects segments with energy peaks and those presenting high complexity, that is, segments with a wide frequency spectrum. The steps of correcting errors and discarding values that fall outside of reasonable ranges are essential for achieving reliable measurements.

5.2 Stress reactivity

The nature of the stressor, although not studied in this work, appears to significantly influence stress reactivity. In the study by Hu *et al.* [23], individuals with depression exhibited hyperreactivity when subjected to a stressor in the form of an interview, whereas hyporeactivity was observed during a cognitively challenging task. It was suggested by the authors that individuals suffering from depression might demonstrate reduced motivation to achieve positive outcomes in tasks requiring cognitive effort. Conversely, their response to emotional distress experienced in daily life, such as that induced by interviews, may be heightened. Studies reviewed by Kibler and Ma [24] also demonstrated an increase in reactivity during speech tasks. Based on this hypothesis, the hyporeactivity observed in both the ST and the TMT was anticipated. Furthermore, as noted by Salomon *et al.* [25], diminished cardiovascular response and compromised recovery in individuals with current major depression could be influenced by their mood state. This primarily indicates deficiencies in motivation rather than physiological abnormalities within the cardiovascular system.

Although the distributions of the metrics differ between patient and control groups, the variability between subjects is substantial. Therefore, establishing thresholds to determine whether the reactivity corresponds to a healthy or diseased nervous system does not appear to be the optimal

approach. Instead, the trend of reactivity should be examined in subsequent measurements. An increase in reactivity during treatment would suggest an improvement in mental health. To evaluate the feasibility and effectiveness of this follow-up, a longitudinal study is required. Specifically, it is crucial to investigate the effects of performing stressor tests recurrently. As patients become accustomed to these tasks, they may exhibit decreased interest in optimal performance, resulting in reduced stress levels. Additionally, an enhanced ability to complete tasks may contribute to reduced stress. Investigating the variations in habituation between individuals with and without depression could be a compelling area of research. This is particularly pertinent given the theory that decreased engagement in cognitively demanding activities underlies the blunted stress response observed in individuals with depression [23]. Furthermore, Brindle *et al.* [26] suggested that stress reactivity is mediated by stress experience rather than stress exposure. Including test scores as an additional metric could provide valuable insights into the patients' commitment to the tasks.

The effects of anxiety are difficult to remove, and the comorbidity of anxiety and depression is frequent. Numerous studies have observed the inclusion of patients with both depression and anxiety, complicating the attribution of effects solely to depression. Instead, these effects may be attributed to the broader spectrum of anxiety and depression [27]. Several reports in the literature have not identified significant differences when comparing groups with depression and anxiety, either in stress reactivity [23, 28] or in basal levels [28]. Conversely, other studies have yielded different findings. For instance, [29] reported that HF power levels were elevated in patients with comorbid MDD and GAD compared to those with only GAD. Additionally, an investigation that examined skin conductance, skin temperature, pupil diameter, and heart rate in participants exposed to unpleasant, pleasant, and neutral images indicated that anxiety and depression influence autonomic output reactivity differently. This distinction may aid in differentiating individuals with anxiety from those with depression [30]. Furthermore, in [31], individuals with generalized anxiety disorder exhibited reduced reactivity compared to the control group,

while those with social anxiety disorder demonstrated increased reactivity. The incorporation of objective measures, such as HRV, in longitudinal investigations of these conditions could enhance the understanding of their interconnections and distinctions [32].

No comprehensive study has been conducted on the influence of antidepressant medication on these findings. Subjects using tricyclic antidepressants, beta-blockers, and antipsychotics, however, were excluded. Benzodiazepines were not excluded from the analysis, potentially impacting the results. Further investigation is warranted to elucidate the effects of benzodiazepines on stress reactivity and baseline levels and their impact on the study's findings. With respect to non-tricyclic antidepressants, Hu *et al.* [23] demonstrated that although antidepressant medication affects baseline levels, it does not influence stress reactivity.

PPG morphological metrics have been employed to evaluate stress responses, demonstrating distinct differences between individuals with depression and those who are healthy [33]. It is recommended that future research investigates whether similar assessments can be conducted using SCPPG.

5.3 Limitations

Differences in stress reactivity between men and women have been documented. Additionally, variations in women depending on the menstrual cycle phase have been noted [27]. In this study, the groups were matched by sex, though the menstrual phase was not accounted for.

The analysis conducted assumes stationarity within each phase of the protocol, which may not hold true for all cases, particularly in the recovery phase performed immediately after the stressor. Significant differences may exist between the first and last minutes of this phase. Stress accumulation during the tests and relaxation during the basal phase could also impact results. Additional research and supporting evidence are needed to validate these assumptions and understand their implications.

The respiratory signal captured by a band was used for the decomposition

of the modulating signal. In cases where only a smartphone is utilized, this respiratory signal must be extracted from the SCPPG. This process is complex and requires thorough investigation, as discussed in [34].

Ultimately, depression measurements were not performed, preventing a correlation analysis with stress reactivity levels.

6 Conclusions

The results of this study support the effectiveness of SCPPG technology in evaluating the stress response, making it suitable for monitoring depression and anxiety. The PRV metrics show a strong correlation ($r \geq 0.96$, $p < 0.001$) and minimal bias ($\Delta \leq 2\%$) for all metrics excluding RMSSD. While RMSSD presents some bias ($\Delta = 12\%$), it does not conceal the differences between patient and control groups, and the correlation remains significant ($r = 0.96$, $p < 0.001$). Consequently, this technology proves valuable for evaluating both baseline states and stress reactivity. One of the most promising applications of this technology involves investigating it within expanded, longitudinal cohorts, particularly given the ubiquitous integration of smartphones in modern-day society.

Bibliography

- [1] Christopher G. Scully, Jinseok Lee, Joseph Meyer, Alexander M. Gorbach, Domhnall Granquist-Fraser, Yitzhak Mendelson, and Ki H. Chon. Physiological parameter monitoring from optical recordings with a mobile phone. *IEEE Transactions on Biomedical Engineering*, 59(2):303–306, 2 2012.
- [2] Sarah Ali Siddiqui, Yuan Zhang, Zhiquan Feng, and Anton Kos. A Pulse Rate Estimation Algorithm Using PPG and Smartphone Camera. *Journal of Medical Systems*, 40(5), 5 2016.
- [3] Daniel J. Plews, Ben Scott, Marco Altini, Matt Wood, Andrew E. Kilding, and Paul B. Laursen. Comparison of heart-rate-variability recording with smartphone photoplethysmography, polar H7 chest strap, and electrocardiography. *International Journal of Sports Physiology and Performance*, 12(10):1324–1328, 11 2017.
- [4] Ivan Liu, Shiguang Ni, and Kaiping Peng. Happiness at Your Fingertips: Assessing Mental Health with Smartphone Photoplethysmogram-Based Heart Rate Variability Analysis. *Telemedicine and e-Health*, 26(12):1483–1491, 12 2020.
- [5] Yuwei Zhang, Yuan Zhang, Sarah Ali Siddiqui, and Anton Kos. Non-invasive Blood-Glucose Estimation Using Smartphone PPG Signals and Subspace KNN Classifier. *Electrotechnical Review/Elektrotehniski Vestnik*, 86(2):68–74, 2019.
- [6] Lorenzo Dall’Olio, Nico Curti, Daniel Remondini, Yosef Safi Harb, Folkert W. Asselbergs, Gastone Castellani, and Hae Won Uh. Prediction

- of vascular aging based on smartphone acquired PPG signals. *Scientific Reports*, 10(1), 12 2020.
- [7] Muhammad Shabaan, Kaleem Arshad, Muhammad Yaqub, Fung Jinchao, M. Sultan Zia, Girridhar Reddy Boja, Muazzam Iftikhar, Usman Ghani, Loknath Sai Ambati, and Rizwan Munir. Survey: Smartphone-based assessment of cardiovascular diseases using ECG and PPG analysis, 7 2020.
- [8] Diego Cajal, Mar Posadas de Miguel, Concepción de la Cámara, Spyridon Kontaxis, Jesús Lázaro, and Raquel Bailón. Smartphone PPG Validation for a Depression Assessment Protocol. In *2022 10th E-Health and Bioengineering Conference, EHB 2022*, 2022.
- [9] Carolina Varon, Jesus Lazaro, Juan Bolea, Alberto Hernando, Jordi Aguilo, Eduardo Gil, Sabine Van Huffel, and Raquel Bailon. Unconstrained Estimation of HRV Indices after Removing Respiratory Influences from Heart Rate. *IEEE Journal of Biomedical and Health Informatics*, 23(6):2386–2397, 11 2019.
- [10] Jonathan Rottenberg. Cardiac vagal control in depression: A critical analysis. *Biological Psychology*, 74(2):200–211, 2 2007.
- [11] W. Verkruyse, L.O. Svaasand, and J.S. Nelson. Remote plethysmographic imaging using ambient light. *Optics Express*, 16(26):21434–21445, 2008.
- [12] Y. Maeda, M. Sekine, and T. Tamura. The advantages of wearable green reflected photoplethysmography. *Journal of Medical Systems*, 35(5):829–834, 2011.
- [13] Jordi Aguiló, Pau Ferrer-Salvans, Antonio García-Rozo, Antonio Armario, Angel Corbí, Francisco J Cambra, Raquel Bailón, Ana González-Marcos, Gerardo Caja, Sira Aguiló, Raúl López-Antón, Adriana Arza-Valdés, and Jorge M. Garzón-Rey. Project ES3: attempting to quantify and measure the level of stress. *Rev Neurol*, 61(9):405–415, 2015.

- [14] Norman Cliff. Dominance Statistics: Ordinal Analyses to Answer Ordinal Questions. *Psychological Bulletin*, 114(3):494–509, 11 1993.
- [15] Kristen Salomon, April Clift, Mardís Karlsdóttir, and Jonathan Rotenberg. Major Depressive Disorder Is Associated With Attenuated Cardiovascular Reactivity and Impaired Recovery Among Those Free of Cardiovascular Disease. *Health Psychology*, 28(2):157–165, 3 2009.
- [16] Jennifer L. Hocking Schuler and William H. O’Brien. Cardiovascular recovery from stress and hypertension risk factors: A meta-analytic review, 1997.
- [17] Jessica L. Hamilton and Lauren B. Alloy. Atypical reactivity of heart rate variability to stress and depression across development: Systematic review of the literature and directions for future research, 12 2016.
- [18] A. Choi and H. Shin. Photoplethysmography sampling frequency: Pilot assessment of how low can we go to analyze pulse rate variability with reliability? *Physiological Measurement*, 38(3):586–600, 2 2017.
- [19] Szabolcs Béres and László Hejmel. The minimal sampling frequency of the photoplethysmogram for accurate pulse rate variability parameters in healthy volunteers. *Biomedical Signal Processing and Control*, 68, 7 2021.
- [20] Giulio Lovisotto, Henry Turner, Simon Eberz, and Ivan Martinovic. Seeing Red: PPG biometrics using smartphone cameras. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, volume 2020-June, pages 3565–3574. IEEE Computer Society, 6 2020.
- [21] Diego Cajal, David Hernando, Jesús Lázaro, Pablo Laguna, Eduardo Gil, and Raquel Bailón. Effects of Missing Data on Heart Rate Variability Metrics. *Sensors*, 22(15), 2022.
- [22] Andrew H. Kemp, Daniel S. Quintana, Marcus A. Gray, Kim L. Felmingham, Kerri Brown, and Justine M. Gatt. Impact of Depression

- and Antidepressant Treatment on Heart Rate Variability: A Review and Meta-Analysis. *Biological Psychiatry*, 67(11):1067–1074, 6 2010.
- [23] Mandy X. Hu, Femke Lamers, Eco J.C. De Geus, and Brenda W.J.H. Penninx. Differential autonomic nervous system reactivity in depression and anxiety during stress depending on type of stressor. *Psychosomatic Medicine*, 78(5):562–572, 6 2016.
- [24] Jeffrey L Kibler and Mindy Ma. Depressive Symptoms and Cardiovascular Reactivity to Laboratory Behavioral Stress. *International Journal of Behavioral Medicine*, 11(2):81–87, 2004.
- [25] Kristen Salomon, Lauren M. Bylsma, Kristi E. White, Vanessa Panaite, and Jonathan Rottenberg. Is blunted cardiovascular reactivity in depression mood-state dependent? A comparison of major depressive disorder remitted depression and healthy controls. *International Journal of Psychophysiology*, 90(1):50–57, 10 2013.
- [26] Ryan C. Brindle, Annie T. Ginty, and Sarah M. Conklin. Is the association between depression and blunted cardiovascular stress reactions mediated by perceptions of stress? *International Journal of Psychophysiology*, 90(1):66–72, 10 2013.
- [27] Carmen Schiweck, Deborah Piette, Daniel Berckmans, Stephan Claes, and Elske Vrieze. Heart rate and high frequency heart rate variability during stress as biomarker for clinical depression. A systematic review. *Psychological Medicine*, 49(2):200–211, 1 2019.
- [28] Katharina Kircanski, Christian E. Waugh, M. Catalina Camacho, and Ian H. Gotlib. Aberrant parasympathetic stress responsivity in pure and co-occurring major depressive disorder and generalized anxiety disorder. *Journal of Psychopathology and Behavioral Assessment*, 38(1):5–19, 3 2016.
- [29] Stefan G. Hofmann, Stefan M. Schulz, Sanna Heering, Frederick Muench, and Lynn F. Bufka. Psychophysiological correlates of generalized anxiety disorder with or without comorbid depression. *International Journal of Psychophysiology*, 78(1):35–41, 10 2010.

- [30] Lucas De Zorzi, Stéphane Ranfaing, Jacques Honoré, and Henrique Sequeira. Autonomic reactivity to emotion: A marker of sub-clinical anxiety and depression symptoms? *Psychophysiology*, 58(4), 4 2021.
- [31] David F. Tolin, Eric Lee, Hannah C. Levy, Akanksha Das, Liya Mammo, Benjamin W. Katz, and Gretchen J. Diefenbach. Psychophysiological assessment of stress reactivity and recovery in anxiety disorders. *Journal of Anxiety Disorders*, 82, 8 2021.
- [32] Terrie E. Moffitt, Avshalom Caspi, Honalee Harrington, Barry J. Milne, Maria Melchior, David Goldberg, and Richie Poulton. Generalized anxiety disorder and depression: Childhood risk factors in a birth cohort followed to age 32. *Psychological Medicine*, 37(3):441–452, 3 2007.
- [33] Spyridon Kontaxis, Eduardo Gil, Vaidotas Marozas, Jesus Lazaro, Esther Garcia, Mar Posadas-De Miguel, Sara Siddi, Maria Luisa Bernal, Jordi Aguilo, Josep Maria Haro, Concepcion De La Camara, Pablo Laguna, and Raquel Bailon. Photoplethysmographic Waveform Analysis for Autonomic Reactivity Assessment in Depression. *IEEE Transactions on Biomedical Engineering*, 68(4):1273–1281, 4 2021.
- [34] Jesús Lázaro, Yunyoung Nam, Eduardo Gil, Pablo Laguna, and Ki H. Chon. Respiratory rate derived from smartphone-camera-acquired pulse photoplethysmographic signals. *Physiological Measurement*, 36(11):2317–2333, 10 2015.

Early detection of sleep apnea (I)

1 Objectives and motivations

OSA is a disease of high prevalence in the general population, yet it remains often underdiagnosed. The gold standard in clinical practice for its diagnosis and severity assessment is supervised polysomnography, although in-home alternatives have been proposed in recent years to overcome its limitations. Today's ubiquitous presence of wearables may become a powerful screening tool in the general population, and pulse-oximetry-based techniques could be used for early OSA diagnosis. This study employs peripheral oxygen saturation and the pulse-to-pulse interval (PPI) series, derived from PPG, as inputs for diagnosing OSA. Different models are trained to classify between normal and abnormal breathing segments (binary decision) and between normal, apneic, and hypopneic segments (multiclass decision). A novel index, the cyclic variation of the heart rate index (CVHRI), derived from the PPI's spectrum, is computed on the segments containing disturbed breathing, representing the frequency of the events. CVHRI is evaluated as an instrument to distinguish subjects based on whether their AHI is above or below the diagnosis thresholds of 5 and 15. In summary, the study investigates whether patient categorization through the combined analysis of oxygen saturation, PPI, and CVHRI provides an effective, user-friendly, and economical method for home-based OSA screening.

2 Introduction

In [1], a novel CVHR-based OSA screening method was proposed and preliminarily evaluated with recordings from 15 subjects. CVHR was detected from the PPI signal using its Hjorth parameters as input of a bagged-tree model. In addition, a frequency-based metric, CVHRI, was proposed for severity stratification, obtaining a Pearson's correlation (r) of $r = 0.68$ ($p < 0.05$) with AHI.

The SpO₂ signal is added to the model in this study, hypothesizing that it may considerably improve segment classification outcomes given that it provides a different source of information of the apnea-generated hypoxia. In addition, further inclusion of PRV metrics in the frequency domain as predictors is studied. Different combinations of PPI and SpO₂ inputs are used in order to understand their individual contribution.

CVHRI is evaluated with recordings from 96 subjects as a potential metric for stratification of subjects with AHI higher or lower than a threshold, considered 5 or 15, corresponding to OSA diagnosis thresholds with and without symptoms or comorbidities, respectively. The novelties of this work are summarized as follows:

- The use of Hjorth parameters as the only features of SpO₂ and PPI signals for the classification of segments.
- The inclusion of the SpO₂ signal to the model preliminary presented in [1]. A larger dataset of 96 subjects is also used, in comparison with the previous 15-subject dataset.
- The inclusion of PRV metrics as model inputs is also studied.
- The use of a new PPI-derived index, the CVHRI, for stratifying subjects between OSA/non-OSA.

3 Methods

3.1 Dataset

The dataset is composed of 96 subjects (age 44.5 ± 11.4 years, 62 males) suspected to suffer from OSA, who underwent a PSG (Medatec, Brainnet II, Brussels, Belgium) at the sleep laboratory of the University Hospital Leuven (UZ Leuven, Belgium).

All included patients did not suffer from any of the following comorbidities: atrial fibrillation, hypertension, stroke, myocardial infarction, hyperlipidemia, or diabetes. All signals were sampled at 500 Hz, including nasal pressure and oronasal flow (thermistor). Hypnograms are also available. PPG and SpO₂ signals were recorded using a Nonin 8000J sensor at 500 Hz.

One subject was removed from the dataset due to him/her being awake most of the test time, whereas another was removed due to an unreliable nasal pressure signal. Thus, in total, 94 subjects were used. 72% of the subjects had $AHI \geq 5$, while 50% had $AHI \geq 15$, according to the AASM annotation rules [2]. The inclusion of these data sets was approved by the ethical committee of UZ Leuven (S60319) and all patients signed an informed consent.

3.2 Signals for segment classification

Pulse-to-pulse intervals

The PPG signal is processed to obtain pulse event series using the adaptive threshold pulse detector. Then, pulse series are checked using the algorithm described in Chapter 6, correcting both false positives and false negatives. Finally, the PPI signal is obtained by evenly sampling the pulse series at 4 Hz using linear interpolation. The PPI signal is also smoothed using a second-order polynomial fitting with a moving window of 20 seconds (see Figure 45).

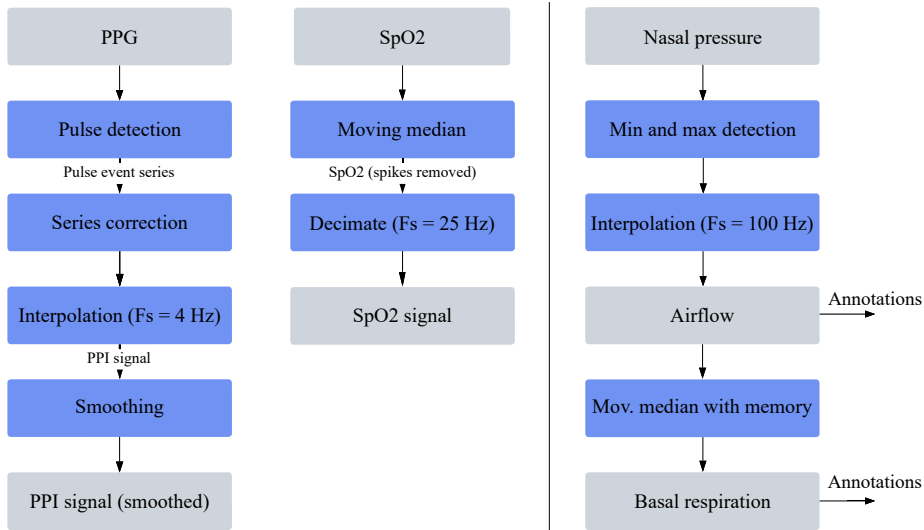


Figure 45: Signal processing flowchart of signals employed in segment classification (left) and annotation (right). Gray boxes represent signals and blue boxes represent processes.

SpO₂

SpO₂ values are quantified in integer units, and the lack of hysteresis provokes large quantization noise. A 3-second median filter is used to reduce noise in the SpO₂ signal. SpO₂ is also decimated to 25 Hz, following the AASM 2012 recommendations [2].

3.3 References for performance evaluation

Airflow

The AASM [2] recommends different sensors to annotate apneas and hypopneas: apneas are proposed to be annotated from the oronasal thermistor, while hypopneas are annotated from nasal pressure. However, oronasal thermistor signals were saturated in most cases, making them unreliable. Thus, following the AASM guidelines for these cases, both apneas and hypopneas are annotated from nasal pressure signals.

First, nasal pressure signals are low-pass filtered at 15 Hz for noise removal, and detrended by a high-pass filter at 0.1 Hz, using 3rd-order Butterworth filters. Airflow is computed from the nasal pressure signal by the algorithm described in [3] by detection of maxima (for positive segments) or minima (for negative segments) between consecutive zero crossings. These minima and maxima are interpolated using piecewise cubic Hermite interpolating polynomials, obtaining a positive and a negative envelope. Finally, the airflow is defined as the difference between the envelopes. The airflow signal is decimated to 100 Hz, following the AASM 2012 recommendations [2].

Basal respiration

Running basal respiration was used as a reference to annotate reductions in the airflow signal. This signal is obtained by an algorithm that computes the median of the airflow signal in 1-minute segments. The result of the sum of the median of each segment, multiplied by a weight of 0.4, plus the median of the previous segment, multiplied by 0.6, is stored. These weights were chosen empirically, looking for the smoothest line that at the same time allowed following the variations in the baseline in a subset of 10 random subjects.

The use of averaging with memory helps to obtain a more accurate basal respiration in signals with the presence of apneas. Without these disruptions, averaging is a simple task, and no weighting is required. However, subjects with apneas present regions with large variations, sometimes composed of a succession of events. Therefore, a compromise must be reached that allows these disruptions to not increase or lower basal respiration to arbitrary values, while allowing the average to follow the changes over time.

The more challenging decision is made in those cases with burst events, where respiration does not return to a basal respiration between events. In these cases, the basal values before and after the burst must be taken into account to make a correct approximation. Thus, once the algorithm obtains a value for each airflow segment, it is run again backwards. Finally, basal respiration is the mean of the forward and backward results. This allows

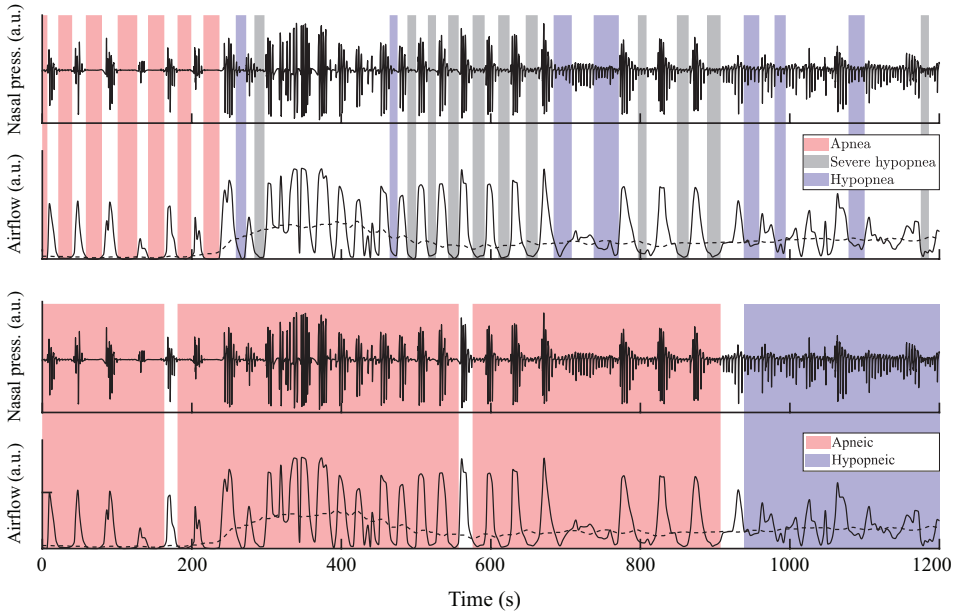


Figure 46: *Top panel: Event-based annotations.* Solid lines represent nasal pressure (top row) and airflow (bottom row). The dashed line represents basal respiration. The onsets and endsets of the annotation correspond to the reductions in airflow and the restorations, respectively. *Bottom panel: Grouped annotations.* Events are grouped in apneic/hypopneic bursts. Note that new bursts are initiated every 8 events, allowing a precise apneic/hypopneic characterization. These annotations are then transformed into segment-based annotations.

for more accurate transitions between normal and disrupted breathing segments.

Annotations

Annotations were made following the AASM 2012 guidelines [2]. Events are labeled as *apnea* if airflow decreases $\geq 90\%$ from basal respiration, during ≥ 10 seconds, while they are labeled as *hypopnea* if the decrease is $\geq 30\%$ during ≥ 10 seconds and there is an associated $\geq 3\%$ desaturation. A third label, *severe hypopnea*, is applied to decreases in airflow $\geq 70\%$ during ≥ 10 seconds, regardless of saturation. These are borderline cases in which airflow does not completely disappear, although the reduction is considerably

greater than in most hypopneas. This third classification is not included within the AASM guidelines, but is an additional definition developed independently by the authors. Hypopneas related to arousals were not annotated. No distinction was made between central and obstructive apnea/hypopnea annotations, although respiratory effort was assessed. 89.7% of apneas/hypopneas were obstructive [4].

Event-based annotations were transformed into segment-based annotations. This step is performed taking into account that the objective is to annotate each segment into *abnormal/normal breathing* (binary decision) or *apneic/hypopneic/normal breathing* (multiclass decision) as a reference for segment classification. Events are grouped in bursts. Bursts are composed of at least two events, separated by a maximum of 180 seconds, and are labeled as apneic or hypopneic depending on the events forming each burst. A burst is labeled as *apneic* if it contains at least one *apnea* or at least half of the events are *severe hypopneas*. Bursts are labeled as *hypopneic* otherwise. The maximum number of grouped events is empirically set to eight, allowing precise *apneic/hypopneic* characterization. An example of event-based annotation grouping in bursts is shown in Figure 46.

Finally, time is divided into 180-second segments with a 150-second overlap. A segment is labeled as *apneic* if it contains at least one *apneic* burst; as *hypopneic* if it contains at least one *hypopneic* burst and no *apneic* bursts; or as *normal breathing* if it does not contain any burst. *Apneic* and *hypopneic* classes are grouped into *abnormal breathing* in the binary case. To address class imbalance, the majority class was randomly downsampled before segment classification. As a result, the binary classification dataset included a total of 53,242 segments, evenly distributed between 26,621 *normal* and 26,621 *abnormal breathing* segments. For the multiclass case, the dataset comprised 25,278 segments, balanced across three categories: 8,426 *apneic*, 8,426 *hypopneic* and 8,426 *normal breathing* segments.

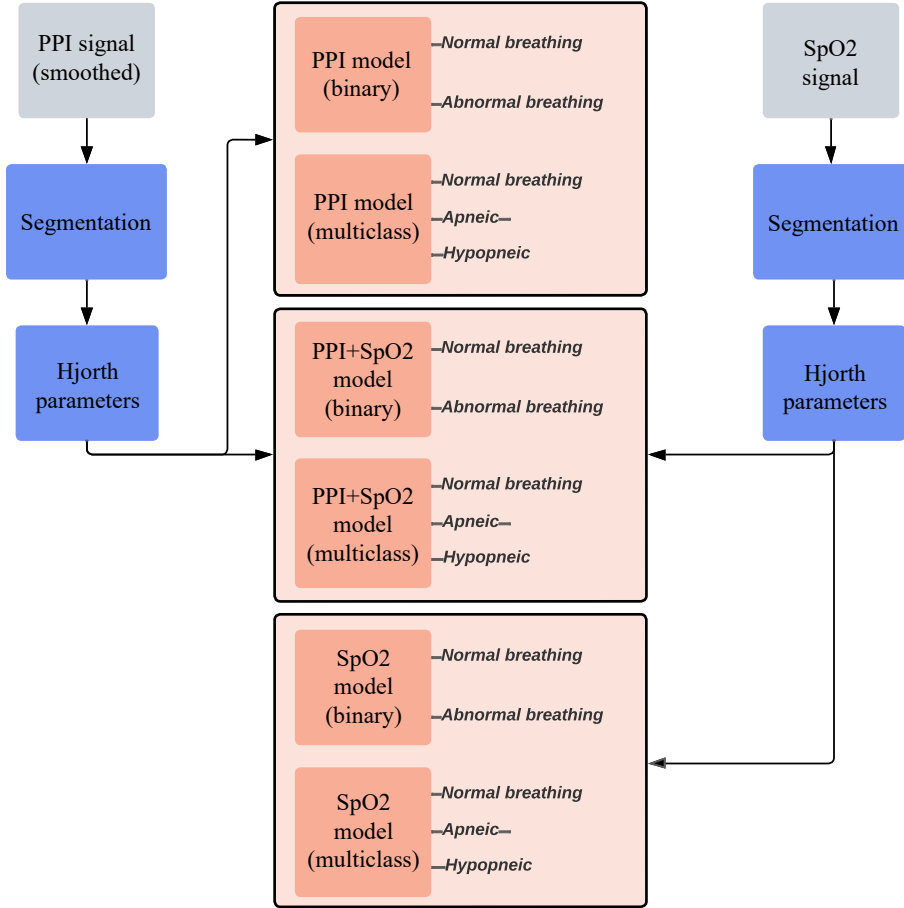


Figure 47: Segment classification models. Gray boxes represent signals, blue boxes represent processes and red boxes represent models.

3.4 Segment classification models

For each segment, the Hjorth parameters \mathcal{H}_0 , \mathcal{H}_1 and \mathcal{H}_2 of the PPI and SpO₂ signals, surrogates of power, dominant frequency, and bandwidth, respectively, were calculated in a sliding window for the PPI and SpO₂ signals (see Chapter 4, Section 2.1). The Hjorth parameters are used as inputs for segment classification. The motivation for using these parameters is two-fold. On the one hand, they are simple and low-cost to compute since they can be estimated from the time domain signal. On the other hand, they are

easily interpretable, being related to the signal energy, dominant frequency, and bandwidth. The original hypothesis was that segments with CVHR patterns would have different Hjorth parameters than segments without CVHR [1]. In particular, it was hypothesized that the PPI in CVHR segments would have a lower Complexity (\mathcal{H}_2) as the bradycardia-tachycardia pattern would mask the normal variability of the heart, causing it to more closely resemble a sinusoid, as shown in [1]. Analogously, desaturations in SpO_2 should affect all parameters, especially Activity (\mathcal{H}_0).

The best model is selected by maximizing the area under the curve (AUC) of the receiver operating characteristics (ROC) curve. A variety of models, including decision trees, discriminant analysis, logistic regression, naive Bayes, support vector machine, nearest neighbor, kernel approximation, ensembles, and neural networks, were evaluated. Ultimately, the bagged trees model demonstrated superior performance compared to the others and was consequently chosen. Different models were created for binary and multiclass decision. Furthermore, for each classification strategy, three models were created, depending on whether they use PPI and SpO_2 (PPI+ SpO_2 model), only PPI (PPI model) and only SpO_2 (SpO_2 model) as inputs (see Figure 47).

3.5 OSA stratification by CVHRI

CVHRI is a metric proposed in [1] to quantify the severity of apnea. The spectrum of PPI is computed using the FFT for each i -th 180-second segment classified as *abnormal breathing* in the binary case, or as *apneic/hypopneic* in the multiclass case.

The frequency of the FFT modulus maxima, F_i^{max} , between 0 and 0.1 Hz is obtained. Then, CVHRI is defined as the sum of the frequencies of the spectrum peaks of each *abnormal breathing/apneic/hypopneic* segment divided by the total number of segments, obtaining a single parameter that characterizes each patient, similarly to AHI.

$$CVHRI = \frac{\sum_{i=1}^{I_{ab}} F_i^{max}}{I_{tot}}, \quad (8.1)$$

where I_{ab} is the total number of *abnormal breathing* segments in the binary case or the sum of *apneic* and *hypopneic* segments in the multiclass case, and I_{tot} is the total number of segments.

Pearson correlation coefficient (r) between CVHRI and AHI is computed for each model. This index was proposed in [1] as an alternative to detect each bradycardia-tachycardia pattern individually, *i.e.*, the CVHR pattern, to be less expensive and more robust. The former is achieved because only one peak per segment must be detected, which is easier in most cases, while the latter benefits from the highly optimized FFT algorithms.

Finally, CVHRI is used for subject stratification. The groups of interest are those clustered by $AHI < 5$ v. $AHI \geq 5$ and $AHI < 15$ v. $AHI \geq 15$, as are the ones used for the diagnosis of OSA [5]. A CVHRI threshold for each subgroup is searched by linear discriminant analysis. Class weights $w(j)$ are computed to deal with the imbalanced data as follows:

$$w(j) = \frac{N}{2N(j)}, \quad (8.2)$$

where N is the total number of patients and $N(j)$ the number of patients corresponding to class j . Train and test groups are selected randomly, splitting the dataset into two halves. 5-fold cross-validation is used to prevent overfitting during training.

3.6 Performance analysis

Segment classification performance is evaluated in terms of accuracy (Acc) precision (P) and recall (R). These metrics have been evaluated for all subjects and also for $AHI < 15$ and $AHI \geq 15$ subsets. A leave-one-subject-out strategy is implemented by summing up the number of false and true events for each left subject.

After segment classification, CVHRI is calculated for every subject and compared to AHI. Pearson's r correlation was calculated between CVHRI and AHI with a significance level of 0.05. The correlation results are also divided into binary and multiclass depending on the segment classification prior to CVHRI computation. In addition, results are computed for all subjects and for $\text{AHI} < 15$ and $\text{AHI} \geq 15$ subsets separately.

Finally, the stratification results were calculated using the AHI clustering ($\text{AHI} < 5$ v. $\text{AHI} \geq 5$ and $\text{AHI} < 15$ v. $\text{AHI} \geq 15$) as reference. Accuracy, positive predictive value (PPV), sensitivity (Se), negative predictive value (NPV), specificity (Sp), AUC of the ROC, and Cohen's Kappa (κ) are reported alongside the best CVHRI threshold for each model that better clusters the subjects.

Binary segment classification

The target here is *normal* and *abnormal breathing* segment classification. A number of true *normal breathing* segments, T_n , false *normal breathing* segments, F_n , true *abnormal breathing* segments, T_{ab} , and false *abnormal breathing* segments, F_{ab} , are obtained, which are quantified by the metrics in Table 19.

Table 19: Binary segment classification metrics.

Class	Measure	Definition
All	Accuracy	$Acc = (T_n + T_{ab}) / (T_n + F_n + T_{ab} + F_{ab})$
Normal breathing	Recall	$R_n = T_n / (T_n + F_{ab})$
	Precision	$P_n = T_n / (T_n + F_n)$
Abnormal breathing	Recall	$R_{ab} = T_{ab} / (T_{ab} + F_n)$
	Precision	$P_{ab} = T_{ab} / (T_{ab} + F_{ab})$

Multiclass segment classification

Here the classes are *normal breathing*, *apneic* or *hypopneic* segments. The segmentation process yields several categories: true *normal breathing* segments, T_n , and false *normal breathing* segments, F_n , which are further split into those originating from *apneic* and *hypopneic* segments ($F_n = F_{n,ap} + F_{n,h}$); true *apneic* segments, T_{ap} , and false *apneic* segments, F_{ap} , manifested from *normal breathing* and *hypopneic* segments ($F_{ap} = F_{ap,n} + F_{ap,h}$); true *hypopneic* segments, T_h , and their false counterpart, F_h , which include contributions from *normal breathing* and *apneic* segments ($F_h = F_{h,n} + F_{h,ap}$). Classification is assessed using the metrics outlined in Table 20.

4 Results

4.1 Segment classification results

The results for binary and multiclass classification are shown in Tables 21 and 22, respectively. Results are given for all subjects together as well as separately in the $AHI < 15$ and $AHI \geq 15$ subgroups.

Table 20: Multiclass segment classification metrics.

Class	Measure	Definition
All	Accuracy	$Acc = (T_n + T_{ap} + T_h) / (T_n + F_n + T_{ap} + F_{ap} + T_h + F_h)$
Normal breathing	Recall	$R_n = T_n / (T_n + F_{ap,n} + F_{h,n})$
	Precision	$P_n = T_n / (T_n + F_n)$
Apneic	Recall	$R_{ap} = T_{ap} / (T_{ap} + F_{n,ap} + F_{h,ap})$
	Precision	$P_{ap} = T_{ap} / (T_{ap} + F_{ap})$
Hypopneic	Recall	$R_h = T_h / (T_h + F_{n,h} + F_{ap,h})$
	Precision	$P_h = T_h / (T_h + F_h)$

4.2 CVHRI correlation with AHI

Binary

Figure 48 shows CVHRI v. AHI scatter plots for each binary segment classification model. A very strong correlation ($r = 0.94$) is found when CVHRI segments are detected with both the PPI+SpO₂ model and the SpO₂ model. No correlation is found when CVHRI segments are detected by the PPI model, except for the group with $\text{AHI} \geq 15$, where a low correlation ($r = 0.37$) is obtained. The correlation is slightly lower when including only the $\text{AHI} \geq 15$ group compared to the correlation of all subjects: 0.91 using the PPI+SpO₂ model and 0.89 using SpO₂. No correlation is found in any case for $\text{AHI} < 15$.

Multiclass

Figure 49 shows CVHRI v. AHI scatter plots for each multiclass segment classification model. The results when using a multiclass classifier prior to CVHRI computation are analogous to the binary case, obtaining slightly lower values. A very strong correlation is found using the PPI+SpO₂ model ($r = 0.91$) and the SpO₂ model ($r = 0.89$), while a low correlation is found when using the PPI model ($r = 0.32$). Correlation is again slightly lower when including only the $\text{AHI} \geq 15$ group in comparison with all subject correlation, being $r = 0.88$ using the PPI+SpO₂ model and $r = 0.86$ using the SpO₂ model, with the exception of the PPI model, which increased its correlation to $r = 0.45$. No correlation is found in any case for $\text{AHI} < 15$.

4.3 OSA stratification by CVHRI

Stratification results after binary and multiclass segment classification are shown in Tables 23 and 24, respectively.

Table 21: Binary segment classification results (%).

Model	Subgroup	Acc	P_n	R_n	P_{ab}	R_{ab}
PPI+SpO ₂	All subjects	85.0	90.9	87.0	73.5	80.5
	AHI<15	87.3	95.2	90.6	38.1	56.0
	AHI≥15	83.0	84.3	81.5	81.8	84.5
SpO ₂	All subjects	86.3	91.3	88.6	76.1	81.2
	AHI<15	89.4	95.0	93.2	44.4	52.6
	AHI≥15	83.6	85.4	81.4	81.9	85.8
PPI	All subjects	60.3	81.1	57.3	39.4	67.5
	AHI<15	57.9	94.7	57.2	12.7	66.0
	AHI≥15	62.5	66.1	57.6	59.5	67.8

Table 22: Multiclass segment classification results (%).

Model	Subgroup	Acc	P_n	R_n	P_{ap}	R_{ap}	P_h	R_h
PPI+SpO ₂	All subjects	73.1	92.9	80.2	65.9	60.7	23.7	49.9
	AHI<15	80.4	96.0	84.5	16.3	26.0	19.6	48.2
	AHI≥15	66.8	87.8	73.5	73.1	63.4	25.7	50.5
SpO ₂	All subjects	71.7	93.9	77.2	78.2	57.5	22.9	63.5
	AHI<15	79.4	96.4	82.6	32.1	16.8	19.3	65.4
	AHI≥15	65.1	89.5	68.8	80.7	60.7	24.8	62.8
PPI	All subjects	44.4	78.4	48.5	24.8	37.1	10.2	28.9
	AHI<15	45.8	93.6	47.3	4.3	40.0	5.3	24.6
	AHI≥15	43.1	62.9	50.5	42.1	36.8	14.4	30.6

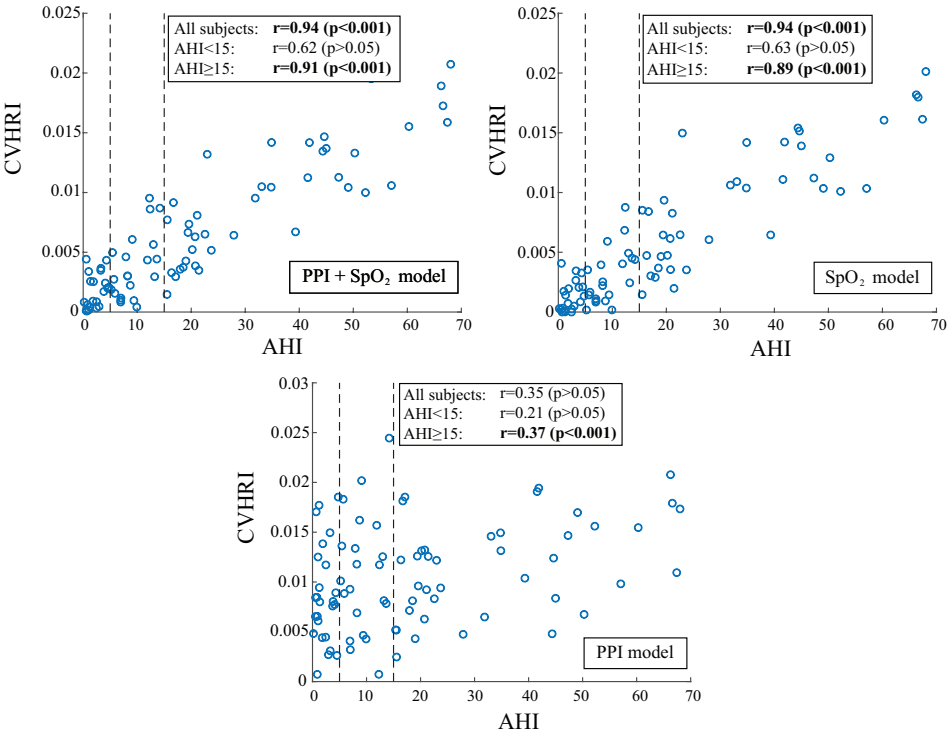


Figure 48: AHI v. CVHRI after binary segment classification. Vertical lines separate AHI < 5, 5 ≤ AHI < 15 and AHI ≥ 15 groups. *Top left:* Segment classification using PPI + SpO₂. *Top right:* Segment classification using SpO₂. *Bottom:* Segment classification using PPI.

Table 23: Stratification after binary segment classification (% , except AUC and κ).

AHI ≥	Model	CVHRI Thresh.	Acc	Se	Sp	PPV	NPV	κ	AUC (Train)	AUC (Test)
5	PPI+SpO ₂	4.0×10 ⁻³	77.3	87.5	75.0	43.8	96.4	0.45	0.88	0.89
	SpO ₂	4.0×10 ⁻³	72.7	87.5	69.4	38.9	96.2	0.38	0.87	0.89
	PPI	1.1×10 ⁻³	51.1	66.7	47.4	23.1	85.7	0.08	0.72	0.56
15	PPI+SpO ₂	5.4×10 ⁻³	79.6	82.4	77.8	70.0	87.6	0.58	0.93	0.90
	SpO ₂	6.3×10 ⁻³	79.6	88.2	74.1	68.2	90.9	0.59	0.96	0.91
	PPI	1.2×10 ⁻³	59.6	57.1	61.5	54.6	64.0	0.19	0.67	0.57

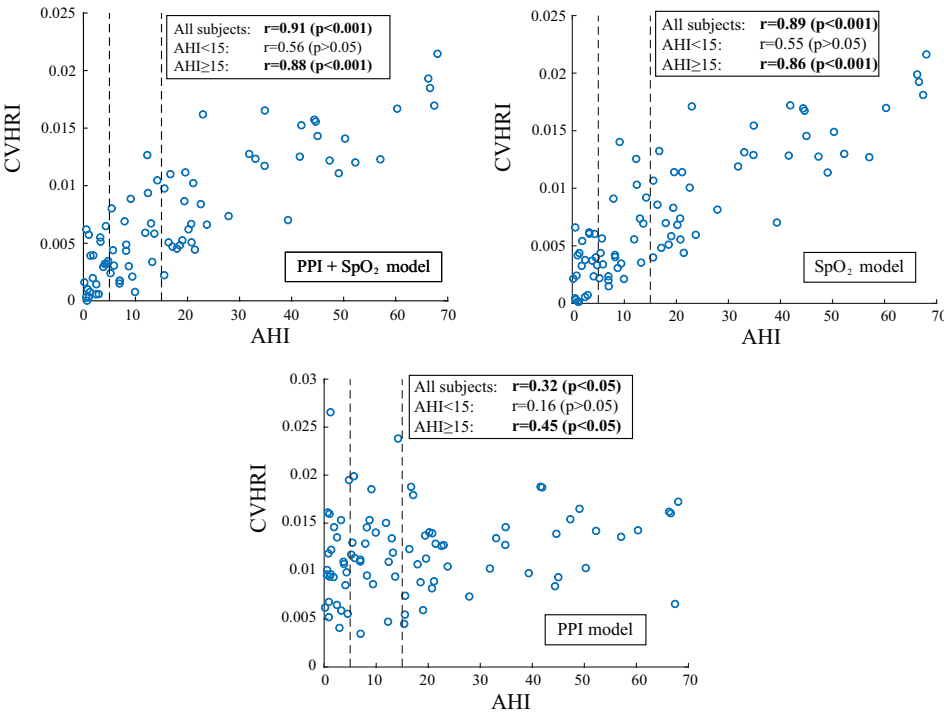


Figure 49: AHI v. CVHRI after multiclass segment classification. Vertical lines separate AHI < 5, 5 ≤ AHI < 15 and AHI ≥ 15 groups. *Top left:* Segment classification using PPI + SpO₂. *Top right:* Segment classification using SpO₂. *Bottom:* Segment classification using PPI.

Table 24: Stratification after multiclass segment classification (% , except AUC and κ).

AHI ≥	Model	CVHRI Thresh.	Acc	Se	Sp	PPV	NPV	κ	AUC (Train)	AUC (Test)
5	PPI+SpO ₂	5.6×10^{-3}	75.0	75.0	75.0	40.0	93.1	0.37	0.88	0.86
	SpO ₂	4.0×10^{-3}	72.7	75.0	72.2	37.5	92.9	0.34	0.86	0.86
	PPI	1.2×10^{-3}	53.2	55.6	52.6	21.7	83.3	0.05	0.71	0.52
15	PPI+SpO ₂	7.8×10^{-3}	79.6	88.2	74.1	68.2	90.9	0.59	0.92	0.88
	SpO ₂	7.8×10^{-3}	77.3	82.4	74.1	66.7	87.0	0.54	0.93	0.90
	PPI	1.4×10^{-3}	46.8	57.1	38.5	42.9	52.6	0.04	0.64	0.51

5 Discussion

The present work has found sufficient evidence to support the use of spectral features, extracted by Hjorth parameters, as models for OSA screening based on oximetry systems. These models stand out for their low computational cost, —linearly proportional to the length P of the segment, in contrast to FFT, with a computational cost proportional to $P \log_2(P)$, or wavelet transform, also proportional to P but with a greater number of operations [6]—, being suitable for built-in wearable applications. Moreover, CVHRI computed on not *normal breathing* classified segments has demonstrated to be strongly correlated to AHI, implying that it could be a proper AHI surrogate when the airflow is not available, especially in moderate-to-severe cases ($\text{AHI} \geq 15$).

However, the limitations that apply to AHI as a diagnosis and severity stand-alone index should also be attributed to CVHRI, making it just as valuable as a surrogate for AHI. As future research, it might be interesting to include patients with comorbidities. As CVHR amplitude is a mortality risk predictor [7], a new index could be included as the amplitude of the peak between 0 and 0.1 Hz of the PPI spectra used in CVHRI computation.

5.1 Segment classification

Both PPI+SpO₂ and SpO₂ models perform similarly, obtaining 85.0% and 86.3% accuracy when including all subjects, respectively. The differences are high compared to the PPI model, which obtained a 60.3% accuracy. Both PPI+SpO₂ and SpO₂ models performed slightly better in the $\text{AHI} < 15$ group, obtaining 87.3% and 89.4% accuracy, in comparison with the $\text{AHI} \geq 15$ group (83.0% and 83.6% accuracy). Despite the better accuracy, precision and recall are uneven between classes in the $\text{AHI} < 15$ group, *e.g.*, 95.2% (*normal* class) against 38.1% (*abnormal breathing* class) precision and 90.6% (*normal* class) against 56.0% (*abnormal breathing* class) recall in the PPI+SpO₂ model; being even in the $\text{AHI} \geq 15$ group, with $\geq 81.5\%$ in all cases.

The multiclass results follow trends similar to those in the binary case.

PPI+SpO₂ and SpO₂ models showed the best performance, with 73.1% and 71.7% accuracy, respectively, whereas the PPI model obtained 44.4% accuracy. Both PPI+SpO₂ and SpO₂ models performed better in the AHI < 15 group as well, obtaining 80.4% and 79.4% against 66.8% and 65.1% accuracy in the AHI ≥ 15 group. Precision and recall are also uneven in the AHI < 15 group in comparison with the AHI ≥ 15 group. Most errors are confusions between *apneic* and *hypopneic breathing*: *e.g.*, in the PPI+SpO₂ model, *normal breathing* precision and recall were 92.9% and 80.2%, while for *apneic/hypopneic breathing* they were 65.9/23.7% precision and 60.7/49.9% recall. These differences are accentuated in the AHI < 15 group: *e.g.*, in the PPI+SpO₂ model, *normal breathing* precision and recall were 96.0% and 84.5%, while for *apneic/hypopneic breathing* they were 16.3/19.6% precision and 26.0/48.2% recall.

According to the results, most of the predictive capacity of the models relies on the SpO₂ signal, taking into account that PPI+SpO₂ models and SpO₂ models perform identically, in contrast with the poor performance of the PPI models. However, correlation between CVHRI and AHI was slightly higher using PPI+SpO₂ models. The inclusion of SpO₂, evaluated in a higher number of subjects, has largely improved the results of the preliminary work [1]. It is possible that PPI could be useful to detect arousal-related hypopneas as [8, 9] suggests.

The overall segment classification performance worsens in the multiclass case relative to the binary case, although accuracies remain high (73.1%). Multiclass classifiers may be useful in future research, especially when including comorbidities, although nowadays there is no distinction in OSA treatment whether there is an apneic or hypopneic predominance. *Severe*

Table 25: Multiclass segment classification results (%). Comparison between using/omitting *severe hypopnea* label using the PPI+SpO₂ model.

<i>Severe hypopnea</i>	<i>Acc</i>	<i>P_n</i>	<i>R_n</i>	<i>P_{ap}</i>	<i>R_{ap}</i>	<i>P_h</i>	<i>R_h</i>
Omitting	73.0	92.9	80.2	65.9	60.7	23.7	49.9
Using	73.1	92.9	80.2	65.9	60.7	23.7	49.9

hypopneas annotations (airflow reduction higher than 70% but lower than 90% related to desaturation) were introduced to improve *apneic* segment detections. 90% reduction of the airflow is an arbitrary threshold that attempts to operationalize the requirement of “absent or nearly absent airflow” [10]. This way, it was observed that borderline events labeled as *hypopneas* have comparable PPI and SpO₂ response to that of apneas rather than < 70% airflow reduction hypopneas. Nevertheless, analyzing the segment classification results, virtually no differences were found when the label *severe hypopnea* was omitted, *i.e.*, strictly following the AASM rules, as it is shown in Table 25. This comparison has been computed using the best classification model (PPI+SpO₂). The utility of this label for stratification is discussed in Section 5.3.

Previous studies have classified apneic events based on SpO₂ and PPI. The approach of this work is different. In this case, there is no detection of apneic/hypopneic events, but rather a classification of segments, which subsequently allows CVHRI to be calculated. To the authors’ knowledge, there are no studies in which segments are classified, which is one of the main novelties of this study. The different approaches make the results not directly comparable with other works. As a reference, it is worth mentioning the event classification results of Lazazzera *et al.* [4], which obtained a 75.1% accuracy on the same database and using the same input signals. This result was obtained for multiclass classification, so it should be compared with the 73.1% (see Table 22) of this work. Deviaene *et al.* [11] obtained an accuracy of 83.4% using SpO₂ and PPG features, in a database with 102 subjects, also recorded at UZ Leuven. In this study, they reached the same conclusion that SpO₂ models outperform PPG models, obtaining an accuracy of 82.2% with the SpO₂ model. The authors concluded that it is better to use both inputs if available [11]. In a recent study by Huttunen *et al.* [12], the authors compare different combinations of signals used as inputs of a deep learning model that is able to simultaneously detect respiratory events and classify sleep stages. The authors compare three models: the first using PPG and SpO₂, the second adding the nasal pressure, and the third using SpO₂, nasal pressure, and the electroencephalogram. Interestingly, the three obtain

virtually the same results in estimating AHI, supporting the use of pulse oximeters in OSA screening without additional sensors.

5.2 CVHRI correlation with AHI

Best correlation between CVHRI and AHI was found in the $AHI \geq 15$ subgroup. This may appear counter-intuitive observing segment classification accuracies, which are higher in the $AHI < 15$ subgroup. However, the *abnormal* class (or *apneic* + *hypopneic* classes in the multiclass model) is better detected in the $AHI \geq 15$ subgroup, according to precision and recall results, probably due to the fact that those are the cases with the most clear respiratory event bursts. As CVHRI is measured only in abnormal (or *apneic* + *hypopneic*) segments, it is not as reliable in the $AHI < 15$ subgroup as in the $AHI \geq 15$ subgroup.

5.3 OSA stratification by CVHRI

Analogously to the previous result sections, stratification accuracy decreases substantially when using the PPI segment classification model. Similar trends follow after multiclass decision. Established thresholds are a reliable tool for OSA diagnosis after PPI+SpO₂ and SpO₂ models, the first being slightly more accurate. Moreover, its use for screening purposes is supported by high negative predictive values. Bad results after PPI models were expected taking into account segment classification results. Binary models

Table 26: Comparison with other studies (% , except AUC and κ).

AHI \geq	Model	Acc	Se	Sp	PPV	NPV	κ	AUC (Train)	AUC (Test)
5	PPI+SpO ₂ binary	77	88	75	44	96	0.45	0.88	0.89
	Romem <i>et al.</i> [13]	-	80	86	93	68	0.67	0.91	-
	Fassbender <i>et al.</i> [14]	-	100	44	62	100	0.43	0.93	-
15	PPI+SpO ₂ binary	80	82	78	70	88	0.59	0.93	0.90
	Romem <i>et al.</i> [13]	-	70	91	80	85	0.71	0.90	-
	Fassbender <i>et al.</i> [14]	-	92	77	60	96	0.59	0.95	-

obtained slightly better results in comparison to multiclass models. This, in addition to the higher complexity of multiclass models, suggests that binary models should be used for stratification purposes.

The results are comparable with other research that uses PPG-derived metrics for OSA diagnosis (see Table 26), such as [13, 14], although none of these works used train-test splits nor cross-validation. Also, higher scores in [14] could be explained since airflow information was added to the model by using a nasal cannula. A comparison with [15] cannot be directly done as different groups were used ($5 \leq \text{AHI} < 15$, $15 \leq \text{AHI} < 30$ and $\text{AHI} \geq 30$), obtaining κ values ranging from 0.49 to 0.79. It is reasonable to assume that these results are in the same order as in [13], since they used the same proprietary algorithm (Morpheus Ox. WideMed Ltd, Herziliya, Israel). The same AHI groups were used in [16]. In this work, AHI was estimated from SpO_2 using an artificial neural network. Estimated AHI classified patients in the mentioned groups with 90.9% accuracy.

The inclusion of the label *severe hypopnea* also deserves discussion at this point. This label applies to events with an airflow reduction $\geq 70\%$ for ≥ 10 seconds, regardless of desaturation. Therefore, by omitting this label, events with an airflow reduction between 70% and 90% —at which point they are classified as apneas regardless of desaturations— need to be linked to desaturation to be scored. Thus, there are events previously annotated as *severe hypopnea* that may change to *hypopnea* and events where annotations may be removed. Stratification results did not change substantially by omitting the *severe hypopnea* label. Only two errors arose after binary classification (one subject with $\text{AHI} < 5$ was stratified as with an $\text{AHI} \geq 5$ and the other with $\text{AHI} < 15$ was stratified as with an $\text{AHI} \geq 15$) and one error after multiclass classification (one subject with $\text{AHI} < 15$ was stratified as with an $\text{AHI} \geq 15$) with respect to the results including the *severe hypopnea* label. Considering that the dataset consisted of 94 subjects, the increase in error was 2.1% after binary classification and 1.1% after multiclass classification. Therefore, no large differences were obtained to support the need for the label, although the type of error, *i.e.*, false positive in all cases, may be relevant in a screening tool.

5.4 Limitations

Events related to arousals were not considered in this study. The only annotated events were apneas and desaturation-related hypopneas. This decision was made taking into account that the models rely on CVHR pattern detection and that PPI and SpO₂ were the only signals used. First, CVHR pattern is not present in events not associated with desaturations [17]. Secondly, although decreases in the amplitude fluctuations of PPG (DAP) can be used to assess arousals to some extent [18], its feasibility has not been demonstrated from PPI's Hjorth parameters. Also, arousal assessment using DAPs has the limitation that some of the DAPs are not related to apneic arousals [18]. In any case, thresholds for OSA stratification should be set observing oxygen desaturation index thresholds, rather than AHI's, if arousals are not included [19]. In addition, no distinction was made between central, obstructive, and mixed apneas. The justification is rather similar to that for arousal-related hypopneas, as respiratory effort is not available with limited channel motorization.

Another limitation was introduced by the saturation of the oronasal thermistor signals that led to the use of the alternative oronasal pressure for the assessment of airflow. However, the use of an oronasal pressure sensor instead of a nasal pressure sensor may be considered a halfway solution between recommended and alternative, since nasal pressure is criticized because the signal may show a decrease in amplitude during mouth breathing [10].

The device used for the input signals is a commercial pulse oximeter. Since this work has focused on a screening tool, with the possibility of being used in at-home solutions with wearables, it is possible that the signals available may be of lower quality. Tests should be performed to calculate performance metrics for each case.

Finally, the segment classification models used in this work were designed to detect respiratory event bursts that lead to a CVHR pattern, rather than isolated events. As CVHR is mediated by the parasympathetic system [20], it cannot be detected in patients with autonomic nervous system impairments,

such as autonomic neuropathy, multiple system atrophy, or Guillain-Barre syndrome.

6 Frequency-domain PRV metrics as predictors

The inclusion of PRV metrics from the frequency domain in the study for the detection of apneic segments is justified by the known relationship of OSA with sympathetic overactivity. For this reason, HRV has previously been used as a method to assess cardiac autonomic changes during sleep [21]. To find out whether the inclusion of these metrics can provide even improved results, the best resulting model (PPI+SpO₂ model) was taken as a starting point and power in the low frequency band (P_{LF}), in the high frequency band (P_{HF}) and P_{LF}/P_{HF} ratio were added to the inputs. These features were calculated in the same segments as the other models to make the results comparable. Both P_{LF} and P_{HF} were computed by trapezoidal integration of the power spectral density estimate obtained by periodogram within the classic windows, *i.e.*, 0.04-0.15 Hz for the low frequency and 0.15-0.4 Hz for the high frequency.

To facilitate comparison with the PPI+SpO₂ model, both model results are shown including and excluding the PRV metrics (see Tables 27 and 28). Variations in predictor output are minimal, demonstrating that the inclusion of these metrics does not significantly improve the model. For simplicity, results related to the subsequent computation of CVHRI and its ability to predict OSA are not shown since the variation in the outcome is as imperceptible as for segment classification.

It should be noted that the use of the classic high-frequency band has been criticized [22]. It is known that respiration affects the boundaries of this autonomic component; therefore, it should be studied whether the inclusion of respiratory frequency information in the PRV analysis would allow a more accurate classification.

Table 27: Binary segment classification results including frequency-domain PRV metrics as predictors (%).

Model	Subgroup	Acc	P_n	R_n	P_{ab}	R_{ab}
PPI+SpO ₂	All subjects	85.0	90.9	87.0	73.5	80.5
	AHI<15	87.3	95.2	90.6	38.1	56.0
	AHI≥15	83.0	84.3	81.5	81.8	84.5
PPI+SpO ₂ +PRV	All subjects	85.0	90.7	87.2	73.7	80.0
	AHI<15	87.3	95.1	90.6	37.8	55.2
	AHI≥15	83.0	84.0	82.1	82.1	84.0

Table 28: Multiclass segment classification results including frequency-domain PRV metrics as predictors (%).

Model	Subgroup	Acc	P_n	R_n	P_{ap}	R_{ap}	P_h	R_h
PPI+SpO ₂	All subjects	73.1	92.9	80.2	65.9	60.7	23.7	49.9
	AHI<15	80.4	96.0	84.5	16.3	25.9	19.6	48.2
	AHI≥15	66.8	87.8	73.5	73.1	63.4	25.7	50.5
PPI+SpO ₂ +PRV	All subjects	73.1	92.7	80.5	66.1	60.9	23.1	47.8
	AHI<15	80.4	95.9	84.6	17.0	27.6	19.2	46.3
	AHI≥15	66.9	87.5	74.2	73.4	63.5	25.0	48.4

7 Conclusion

A classifier for during-sleep breathing segments has been presented. This classifier exploits the differences in oscillatory pattern characteristics of the SpO₂ and PPI signals by using the Hjorth parameters as features. This approach obtained 86.3% accuracy in the binary (*normal-abnormal breathing*) decision, and 73.1% accuracy in the multiclass (*normal breathing-apneic-hypopneic*) decision. A novel index, CVHRI, has been computed in *normal breathing* segments after segment classification. This index has been shown to be strongly correlated with AHI both after binary ($r = 0.94$, $p < 0.001$) and multiclass ($r = 0.91$, $p < 0.001$) segment classification. A better performance has been found in subjects with $\text{AHI} \geq 15$ rather than in the $\text{AHI} < 15$ subgroup. In addition, CVHRI has been used to stratify the $\text{AHI} \geq 5$ and $\text{AHI} \geq 15$ subgroups, resulting in 77.3% and 79.6% accuracy, respectively. These results suggest that the presented methods provide value for OSA limited-channel screening, allowing monitoring with wearables at home.

Bibliography

- [1] Diego Cajal, Eduardo Gil, Pablo Laguna, Carolina Varon, Dries Testelmans, Bertien Buyse, Chris Jensen, Rohan Hoare, Raquel Bailón, and Jesús Lázaro. Sleep apnea severity stratification by an FFT-based PPG-derived index. In *2022 12th Conference of the European Study Group on Cardiovascular Oscillations, ESGCO 2022*. Institute of Electrical and Electronics Engineers Inc., 2022.
- [2] RB Berry, R Brooks, CE Gamaldo, Harding SM, Marcus CL, Vaughn BV, and Tangredi MM. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2. *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, 2012.
- [3] Jesús Lázaro, Natasa Reljin, Raquel Bailón, Eduardo Gil, Yeonsik Noh, Pablo Laguna, and Ki H. Chon. Electrocardiogram Derived Respiration for Tracking Changes in Tidal Volume from a Wearable Armband. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 596–599, 2020.
- [4] Remo Lazazzera, Margot Deviaene, Carolina Varon, Bertien Buyse, Dries Testelmans, Pablo Laguna, Eduardo Gil, and Guy Carrault. Detection and Classification of Sleep Apnea and Hypopnea Using PPG and SpO₂ Signals. *IEEE Transactions on Biomedical Engineering*, 68(5):1496–1506, 5 2021.
- [5] AASM. International classification of sleep disorders. *Darien, IL: American Academy of Sleep Medicine*, 3rd edition, 2014.

- [6] João Paulo R.R. Leite and Robson L. Moreno. Heartbeat classification with low computational cost using Hjorth parameters. *IET Signal Processing*, 12(4):431–438, 6 2018.
- [7] Junichiro Hayano, Fumihiko Yasuma, Eiichi Watanabe, Robert M. Carney, Phyllis K. Stein, James A. Blumenthal, Petros Arsenos, Konstantinos A. Gatzoulis, Hiroshi Takahashi, Hideki Ishii, Ken Kiyono, Yoshiharu Yamamoto, Yutaka Yoshida, Emi Yuda, and Itsuo Kodama. Blunted cyclic variation of heart rate predicts mortality risk in post-myocardial infarction, end-stage renal disease, and chronic heart failure patients. *Europace*, 19(8):1392–1400, 8 2017.
- [8] Eduardo Gil, Martín Mendez, José María Vergara, Sergio Cerutti, Anna Maria Bianchi, and Pablo Laguna. Discrimination of sleep-apnea-related decreases in the amplitude fluctuations of ppg signal in children by HRV analysis. *IEEE Transactions on Biomedical Engineering*, 56(4):1005–1014, 4 2009.
- [9] Jesús Lázaro, Eduardo Gil, José María Vergara, and Pablo Laguna. Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children. *IEEE Journal of Biomedical and Health Informatics*, 18(1):240–246, 2014.
- [10] Richard B. Berry, Rohit Budhiraja, Daniel J. Gottlieb, David Gozal, Conrad Iber, Vishesh K. Kapur, Carole L. Marcus, Reena Mehra, Sairam Parthasarathy, Stuart F. Quan, Susan Redline, Kingman P. Strohl, Sally L. Davidson Ward, and Michelle M. Tangredi. Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events. *Journal of Clinical Sleep Medicine*, 8(5):597–619, 2012.
- [11] Margot Deviaene, Jesus Lazaro, Dorien Huysmans, Dries Testelmans, Bertien Buyse, Sabine Van Huffel, and Carolina Varon. Sleep Apnea Detection Using Pulse Photoplethysmography. In *Computing in Cardiology*, volume 2018-September. IEEE Computer Society, 9 2018.

- [12] Riku Huttunen, Timo Leppanen, Brett Duce, Erna S. Arnardottir, Sami Nikkonen, Sami Myllymaa, Juha Toyras, and Henri Korkalainen. A Comparison of Signal Combinations for Deep Learning-Based Simultaneous Sleep Staging and Respiratory Event Detection. *IEEE Transactions on Biomedical Engineering*, 70(5):1704–1714, 5 2023.
- [13] Ayal Romem, Anat Romem, Dafna Koldobskiy, and Steven M. Scharf. Diagnosis of obstructive sleep apnea using pulse oximeter derived photoplethysmographic signals. *Journal of Clinical Sleep Medicine*, 10(3):285–290, 2014.
- [14] Philipp Faßbender, Ali Haddad, Silja Bürgener, and Jürgen Peters. Validation of a photoplethysmography device for detection of obstructive sleep apnea in the perioperative setting. *Journal of Clinical Monitoring and Computing*, 33(2):341–345, 4 2019.
- [15] Yan Li, He Gao, and Yan Ma. Evaluation of pulse oximeter derived photoplethysmographic signals for obstructive sleep apnea diagnosis. *Medicine (United States)*, 96(18), 5 2017.
- [16] Sami Nikkonen, Isaac O. Afara, Timo Leppänen, and Juha Töyräs. Artificial neural network analysis of the oxygen saturation signal enables accurate diagnostics of sleep apnea. *Scientific Reports*, 9(1), 12 2019.
- [17] C Zwillich, T Devlin, D White, N Douglas, J Weil, and R Martin. Bradycardia during sleep apnea. Characteristics and mechanism. *Journal of Clinical Investigation*, 69(6):1286–1292, 1982.
- [18] Eduardo Gil, José María Vergara, and Pablo Laguna. Detection of decreases in the amplitude fluctuation of pulse photoplethysmography signal as indication of obstructive sleep apnea syndrome in children. *Biomedical Signal Processing and Control*, 3(3):267–277, 2008.
- [19] Lalee Varghese, Grace Rebekah, Priya Priya N, Ashwin Oliver, and Regi Kurien. Oxygen desaturation index as alternative parameter in screening patients with severe obstructive sleep apnea. *Sleep Science*, 15(S 01):224–228, 3 2022.

-
- [20] Christian Guilleminault, Roger Winkle, Stuart Connolly, Kenneth Melvin, and Ara Tilkian. CYCLICAL VARIATION OF THE HEART RATE IN SLEEP APNOEA SYNDROME: Mechanisms, and Usefulness of 24 h Electrocardiography as a Screening Technique. *The Lancet*, 323(8369):126–131, 1984.
- [21] Seren Ucak, Hasthi U. Dissanayake, Kate Sutherland, Philip de Chazal, and Peter A. Cistulli. Heart rate variability and obstructive sleep apnea: Current perspectives and novel technologies. *Journal of Sleep Research*, 30(4), 8 2021.
- [22] Alberto Hernando, Jesús Lázaro, Eduardo Gil, Adriana Arza, Jorge Mario Garzón, Raúl López-Antón, Concepción De La Camara, Pablo Laguna, Jordi Aguiló, and Raquel Bailón. Inclusion of Respiratory Frequency Information in Heart Rate Variability Analysis for Stress Assessment. *IEEE Journal of Biomedical and Health Informatics*, 20(4):1016–1025, 7 2016.

Early detection of sleep apnea (II)

In the previous chapter, a method was presented to detect sleep apnea using signals commonly found in wearable devices, such as PPG and SpO₂. However, these signals were obtained from a non-wearable device. This chapter introduces a pilot study employing the same methods, but this time using signals acquired from a wrist-worn wearable device that is used by patients at home with no supervision. The methods for event labeling and their transformation into segment labeling remain identical to those described in the previous chapter. In general, the methodology has been preserved. Since the best results were achieved using the PPI+SpO₂ binary classifier, this was the only model tested. Furthermore, the model has not been retrained with the new signals but used as it was originally trained on the previous database. In this case, the PPI and SpO₂ signals are obtained from the wearable device. CVHRI is also calculated in the same way, using the PPI of the wearable device.

1 Methods

1.1 Dataset

A group of 12 participants (age 49 ± 16 years, including 7 males) was gathered in partnership with the Hospital Universitario Miguel Servet (Zaragoza,



Figure 50: Maxim MAXREFDES103 [1]. A wrist-worn device capable of recording PPG with red, green, and infrared channels.

Spain) as part of the PID2021-126734OB-C21 project, following approval from Aragon's ethics committee (CEICA, PI23/336). During the period of this study, patient enrollment is ongoing. However, a small number of recordings is already available. This dataset includes 8 subjects suspected of suffering from OSA and 4 additional subjects as control cases.

The recordings were performed using a commercially available polygraph, the ApneaLink Air device (ResMed Inc., USA), which provided SpO_2 with a sampling rate of 1 Hz and nasal pressure at 100 Hz signals as a reference; and a Maxim MAXREFDES103 device (Analog Devices Inc., USA) as a wearable, which captured PPG signals at 256 Hz with red, green, and infrared channels. Maxim is a wrist-worn device, as shown in Figure 50.

Reference signals from the Apnealink device were used exclusively to label apneic and hypopneic events. Annotations were automatically computed using the same algorithms outlined in the preceding Chapter. These annotations were then compared to the classifier output using the wearable's signals as input. In addition, the reference was used to calculate the AHI, which was compared to the CVHRI obtained from the wearable.

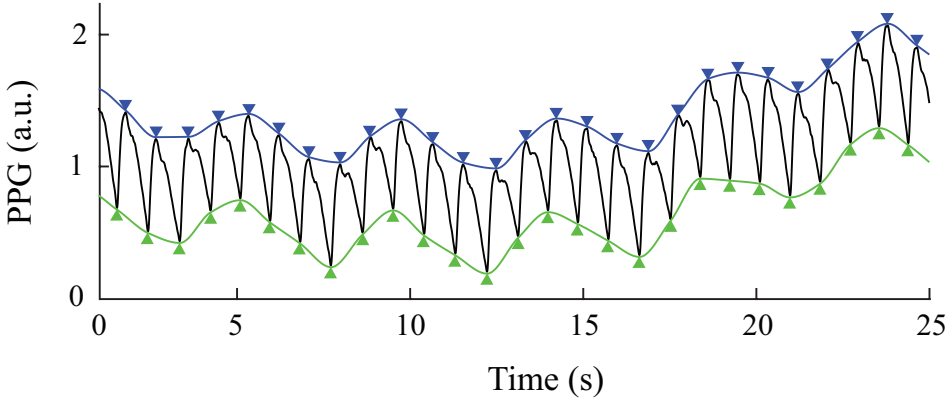


Figure 51: Upper (blue) and lower (green) envelopes of the PPG. Blue triangles indicate $t_O[k]$ points, while green triangles indicate $t_A[k]$ points.

1.2 Peripheral oxygen saturation estimation

The PPI and SpO_2 are estimated from the wearable PPG signal. The PPI is derived from the green channel, following the same methodology described in the previous chapter. To calculate SpO_2 , the R ratio—not to be confused with HRV metrics \mathcal{R} and \mathcal{R}' —has been estimated. This ratio is derived from the red and infrared channels of the PPG signal through a modified version of the formula

$$R = \frac{\frac{AC_{\text{red}}}{DC_{\text{red}}}}{\frac{AC_{\text{infrared}}}{DC_{\text{infrared}}}}, \quad (9.1)$$

where a small value $\epsilon = 0.01$ has been added to the denominator to avoid division by zero:

$$R = \frac{\frac{AC_{\text{red}}}{DC_{\text{red}}}}{\frac{AC_{\text{infrared}}}{DC_{\text{infrared}}} + \epsilon} \quad (9.2)$$

The AC component of each channel is calculated as the difference between the upper and lower envelopes of the PPG signal (see Figure 51). These envelopes are computed by interpolation of $t_A[k]$ and $t_O[k]$ points, respectively (see Chapter 5) at 256 Hz. The DC component is the lower envelope.

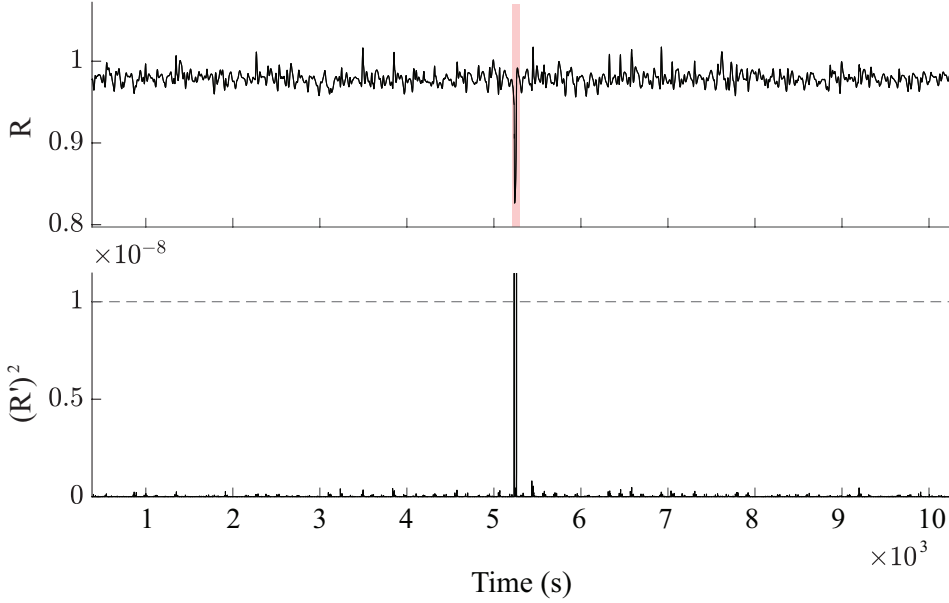


Figure 52: Outlier removal in R *Top:* R signal in black, outlier shaded in red. *Bottom:* Squared derivative of R in solid black. Dashed line corresponds to the fixed threshold at 10^{-8} .

The errors are then removed using a fixed threshold applied to the squared derivative of the signal. This threshold is set at 10^{-8} , effectively eliminating outliers caused primarily by envelope estimation errors (see Figure 52). A second threshold is applied to the R signal to remove values that exceed 4. Subsequently, the signal is low-pass filtered at 0.1 Hz to eliminate rapid variations caused by noise.

According to the wearable documentation, oxygen saturation should be calculated using the formula provided in [2]:

$$\text{SpO}_2 = -16.6(R)^2 + 8.3R + 100. \quad (9.3)$$

However, this formula did not yield satisfactory results with our data. Instead, a subject-specific calibration was performed using the SpO_2 signal of the reference device. From this reference signal, the mean and standard deviation were calculated. Subsequently, the R signal was normalized to obtain a signal with zero mean and unit standard deviation. The normalized

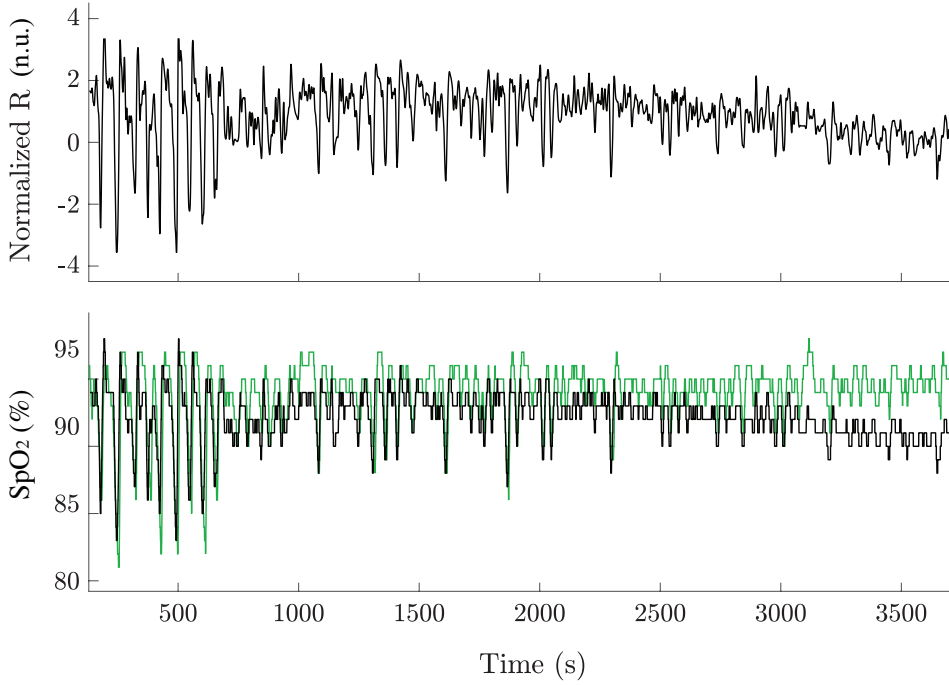


Figure 53: SpO_2 estimation from Maxim. *Top:* Normalized R from Maxim. *Bottom:* SpO_2 from reference in green, estimate from Maxim in black.

signal was then scaled by multiplying it by the standard deviation of the reference and adding its mean. Finally, the signal was resampled at 25 Hz to comply with the recommendations of the AASM [3] and rounded to the nearest integer (see Figure 53).

2 Results

The annotation of events using reference signals resulted in a total of 6,143 *normal breathing* segments and 1,109 *abnormal breathing* segments. The mean AHI was 5.95, with a standard deviation of 6.67. 7 subjects had $\text{AHI} < 5$; 3 subjects had $5 \leq \text{AHI} < 15$; and 2 subjects had $\text{AHI} \geq 15$. Due to the small sample size, the results of the segment classification split into subgroups are not shown as in the previous chapter. Neither stratification analysis is reported. The correlation between CVHRI and AHI is calculated, but the

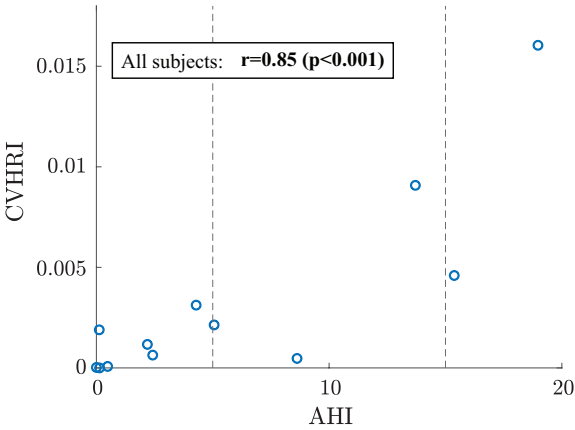


Figure 54: AHI v. CVHRI from a wearable device after binary segment classification. Vertical lines separate $AHI < 5$, $5 \leq AHI < 15$ and $AHI \geq 15$ groups.

results split by subgroups are also not shown.

Following random class balancing, which employs the same method discussed in the preceding chapter, the binary classification results for the segments are consolidated in Table 29. CVHRI correlates with the reference AHI, with a Pearson correlation of $r = 0.85$ ($p < 0.001$). Figure 54 presents a scatter plot of AHI v. CVHRI.

Table 29: Binary segment classification results from a wearable device (%).

Model	Subgroup	Acc	P_n	R_n	P_{ab}	R_{ab}
PPI+SpO ₂	All subjects	65.7	87.8	60.9	43.6	78.2

Table 30: Binary segment classification results using PPI from wearable device and SpO₂ from reference(%).

Model	Subgroup	Acc	P_n	R_n	P_{ab}	R_{ab}
PPI+SpO ₂	All subjects	77.0	88.1	72.1	65.9	84.7

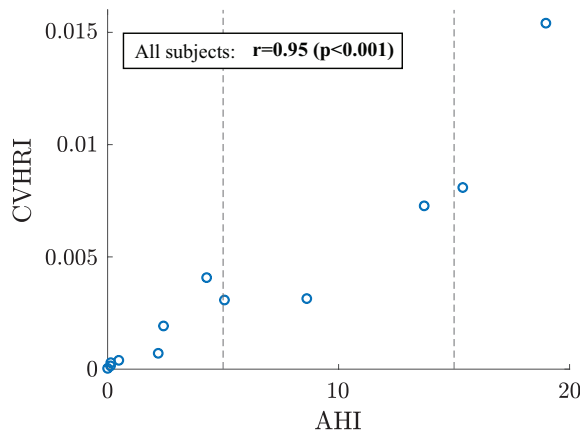


Figure 55: AHI v. CVHRI from a wearable device after binary segment classification. Vertical lines separate $AHI < 5$, $5 \leq AHI < 15$ and $AHI \geq 15$ groups.

To estimate the influence of the estimation SpO_2 based on R, the results of the segment classification are presented using the wearable PPI and the reference SpO_2 as input to the classifier. These results are summarized in Table 30. A new CVHRI is calculated from this classification. Its correlation with AHI becomes $r = 0.95$, $p < 0.001$ (see Figure 55).

3 Discussion

The results obtained using the wearable signals show a degradation compared to those calculated with the polysomnography system in the previous chapter. This was expected for several reasons: a noisier PPG signal prone to artifacts; an SpO_2 estimation derived from this signal; and the use of a pre-trained model designed for other types of signals. Among these factors, the second and third are amenable to improvement.

Estimation of SpO_2 from R could not be performed using the formula provided by the manufacturer. Instead, a subject-specific calibration was used, which has yielded good qualitative results, but requires a quantitative evaluation. Although the need to calibrate a wearable device for each user poses a challenge, it does not render the device unusable. In fact,

it is common practice to request prior calibration from users for other types of measurement, such as blood pressure [4]. However, the approach taken in this experiment could be improved, enabling undoubtedly better classification results, as evidenced by the results obtained using the reference signal SpO_2 . An alternative to this approach would be to retrain the classification model using the Hjorth parameters of R instead of the estimated SpO_2 . Since the parameters reflect the variance, center frequency, and bandwidth of the signal, the operations performed for estimation (*i.e.*, normalization, multiplication by a constant, and addition of a constant) should yield proportional values.

Regarding the classification model, it should be re-trained using the new signals as input. An improvement in results is expected, given the significant differences in the SpO_2 estimate.

The number of subjects should also be increased, as the current sample is limited, with only 2 subjects having an $\text{AHI} \geq 15$. This constrains the analysis that could be performed, excluding the stratification applied in the previous chapter. Moreover, the reduced number of *abnormal breathing* segments means that the classifier's statistics are calculated using 1,109 segments per class, since class balance ensures that each class has the same number of samples as the minority class, in this case, the *abnormal breathing* class. Consequently, the majority of classified segments are excluded from evaluating the classifier's performance. Therefore, more subjects should be recorded, with a particular focus on expanding the subgroup with $\text{AHI} \geq 15$.

Despite this degradation, the results remain promising. The correlation between CVHRI and AHI remains high, decreasing only slightly from 0.94 with the signals from the previous chapter to 0.85 with those from the wearable. This likely enables effective stratification for sleep apnea screening. The drop in correlation is primarily attributed to the decline in classifier accuracy, which decreased from 86.3% to 65.7%. However, this pronounced degradation in accuracy does not correspond to an equivalent percentage drop in the AHI-CVHRI correlation, suggesting that CVHRI may be a robust metric against classifier errors.

Nonetheless, the potential for improving SpO_2 estimation is evident in the results. When the reference SpO_2 is used instead of the wearable-derived SpO_2 , the correlation returns to 0.95, equivalent to that of the previous chapter, and the classifier accuracy improves to 77.0%. While this is not as high as in the previous chapter, it surpasses the accuracy achieved when only wearable-derived inputs are used. This outcome is consistent with the findings of the previous chapter, where Hjorth parameters of SpO_2 were shown to be the main predictors.

4 Conclusions

The results suggest that the method studied in the previous chapter can be applied to wearable devices used at the patient's home with no supervision, delivering good performance. The classifier achieved an accuracy of 65.7%, while the correlation between CVHRI and AHI was $r = 0.85$ ($p < 0.001$). The high correlation indicates that stratification remains feasible. However, the sample size should be increased, particularly for the group with $\text{AHI} \geq 15$, to enable a more comprehensive study. Additionally, efforts should focus on improving the estimation of SpO_2 from the R ratio or directly using the latter as an input. Combined with retraining the model to account for the new characteristics of the signals, these steps are expected to enhance the results further.

Bibliography

- [1] Analog Devices Inc. MAXREFDES103 Health Sensor Platform, 4 2024.
- [2] Guidelines for SpO2 Measurement, 3 2019.
- [3] RB Berry, R Brooks, CE Gamaldo, Harding SM, Marcus CL, Vaughn BV, and Tangredi MM. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2. *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, 2012.
- [4] How to Measure Blood Pressure Using a Galaxy Watch?, 9 2024.

Integrating findings: from theory to practice

1 Primary findings

This thesis elaborates on the objectives set out in Chapter 1, demonstrating a range of techniques that encompass the entire signal processing workflow, from signal capture to the derivation of HRV metrics, specifically tailored to address challenges in wearable environments. These objectives are interconnected, with a shared focus on obtaining reliable metrics to assess the autonomic nervous system. Furthermore, the application of these metrics in the monitoring, screening, and customization of treatments for chronic conditions such as sleep apnea, anxiety, and depression has been thoroughly explored.

One of the key discoveries in the thesis is the development of robust, standardized techniques for examining HRV metrics in the typically fluctuating and noisy environments of wearable devices. This standardization includes the most appropriate methods from the existing literature for filtering, quality assessment, detecting artifacts, identifying events, correcting these events, and calculating variability metrics. In addition to traditional approaches, novel techniques have been developed, including a method to fill artifact-created gaps. Furthermore, a study has been performed to identify the ideal technique for each variability metric, alongside the expected error based on the size of the gap or the percentage of data lost. Guidance on

a general methodology is also provided that emphasizes reducing false positives, which can obstruct further error detection and correction. This is a novel contribution, as the existing literature lacks a comprehensive context of the processing sequence. The thesis approach takes into account both preceding and subsequent processing steps, ensuring that each block strategy is framed with consistency and coherence.

Another fundamental finding is the application of wearable signals in OSA screening. The proposed method is based on the well-known bradycardia-tachycardia pattern produced by apneic events. In this approach, segmentation techniques are employed to differentiate apneic episodes from normal respiration. Additionally, a novel parameter, CVHRI, derived via frequency domain analysis, is utilized due to its correlation with AHI. One of the innovations is the segmentation technique. Here, instead of annotating apneic and hypopneic events separately, entire windows are evaluated during the sleep period, which could include multiple events, including scenarios in which both apneic and hypopneic events exist within the same window. This method provides an alternative method for segmenting sleep time, facilitating analyses such as the frequency analysis mentioned above.

The third primary aspect of the thesis involves developing an application for capturing pulse photoplethysmographic signals using a smartphone camera. Although this is not groundbreaking due to the existence of similar apps, creating a proprietary version has facilitated the use of its own processing methods and enabled control over key recording parameters, including color channel, sampling frequency, filtering, and data format. This app has been used to gather an extensive dataset from patients with anxiety and depression, as well as a control group. Consequently, smartphone PPG has been used for the first time in a study analyzing stress reactivity related to the assessment of mental health disease. This study was carried out in collaboration with professionals and organizations that require such objective metrics to enhance tools for more effective and personalized treatment.

These findings are valuable individually, but they also interrelate in the context of chronic disease monitoring using wearable devices. The devel-

opment of the mobile application serves as a starting point for integrating all the methods discussed into a platform capable of collecting signals from smartphones and connected devices, processing them, and extracting indicators to support the prevention, diagnosis, and treatment of various diseases.

In addition, the pathologies studied are often comorbid and share overlapping symptoms. Advancing research into each of these conditions, whether through the development of new methods or leveraging the power of large-scale data analysis enabled by such devices, can contribute to a better understanding of the others. This may occur through the discovery of common pathways, such as the role of the ANS.

2 Limitations

The study on the degradation of HRV metrics due to data loss does not include an analysis of the effect of false positives. Given the approach taken for the complete processing pipeline, efforts have been made to minimize false positives by applying restrictive criteria during artifact removal stages prior to detection, as well as during the identification of outliers in the detection series. This was done to ensure the availability of a reliable reference for filling in missing data, if necessary. However, the occurrence of some false positives is inevitable, and it would be essential to determine the number or percentage of false positives the proposed methodology can withstand before degradation occurs, as well as to evaluate the extent of the resulting degradation. Another limitation of this study is the inability to accurately fill gaps where the variability is nonmonotonic using Hermite polynomials. The use of these polynomials, instead of other interpolation methods such as cubic splines, which do allow for nonmonotonic interpolation, has been shown to improve results by avoiding the typical overshoot associated with splines. However, alternatives should be explored to achieve a balance between nonmonotonic interpolation and overshoot reduction. This could potentially be addressed by estimating

inflection points for the Hermite polynomials.

The respiratory signal used in the SCPPG validation study was recorded with a non-wearable device. This signal, used for OSP decomposition, should ideally be estimated directly from the SCPPG itself. Furthermore, no data are available to establish a relationship between stress reactivity and levels of anxiety or depression. Differences between males and females have not been analyzed, nor have potential nonstationarities during stressor events been examined.

The primary limitation of the OSA study is the lack of medical professional supervision in event labeling. In addition, electroencephalography was not used, meaning arousals were not considered. Furthermore, no distinction was made between obstructive and central apnea. However, all other AASM guidelines were rigorously followed. This approach aims to minimize annotator subjectivity as much as possible. The second part of the study presents several limitations inherent in a pilot study. The most significant is the size of the dataset, as well as the lack of subjects with severe OSA, with healthy subjects being overrepresented. Another important aspect is the inaccuracy of peripheral saturation estimation derived from the wearable's PPG signal. These two aspects should be addressed in more depth in the full-scale study.

3 Future lines

The described applications have already been implemented and validated in wearable devices, making them ready for immediate practical application. Future studies should focus on aspects such as adherence to monitoring, ease of use, patient comfort, and the use of screening tools by nonpatient populations. Collaboration with medical experts will be essential during this phase to ensure the successful integration of these technologies into clinical and public health contexts. In this regard, the inclusion of numerical metrics, such as those introduced in the study of mental health, poses the additional challenge of being novel in a field predominantly driven by

subjective evaluations. The integration of these metrics is not straightforward and must be designed in a way that is useful to specialists for monitoring and personalizing treatment. This requires careful consideration of how to present the data to ensure that it complements traditional assessments and improves clinical decision making.

The investigation of the best methods for the evaluation of signal quality is still pending. Some of the best methods available in the literature are described in Chapter 4, but, following the approach of this thesis, this processing block must be included in a joint strategy with the rest of the blocks. This is the only way to choose the most appropriate method, which also depends on the final application.

The extension of the study on OSA screening using wearable devices is also a future line. Although preliminary results are promising, they need to be validated in a larger dataset, with particular emphasis on increasing the number of patient-case subjects. This will help ensure the robustness and generalizability of the findings, paving the way for broader clinical adoption. The relationship between CVHRI and OSA-associated symptomatology should also be investigated. As commented above, AHI has often been criticized for not adequately reflecting this relationship. CVHRI is an index that could better capture the effects of OSA on the ANS, as it measures a physiological response rather than airflow reduction.

4 Smarthone app

This section is treated separately due to its significance and relatively advanced stage. As it primarily involves development and implementation rather than novel research, it was not elaborated on in the preceding chapters. Nevertheless, it remains intrinsically linked to the entirety of the work conducted during the Ph.D. The methods and studies discussed previously are meant to be applied to society, moving beyond theoretical concepts to become practical applications. An agreement was reached to consolidate everything examined into one service within the framework of a Proof of

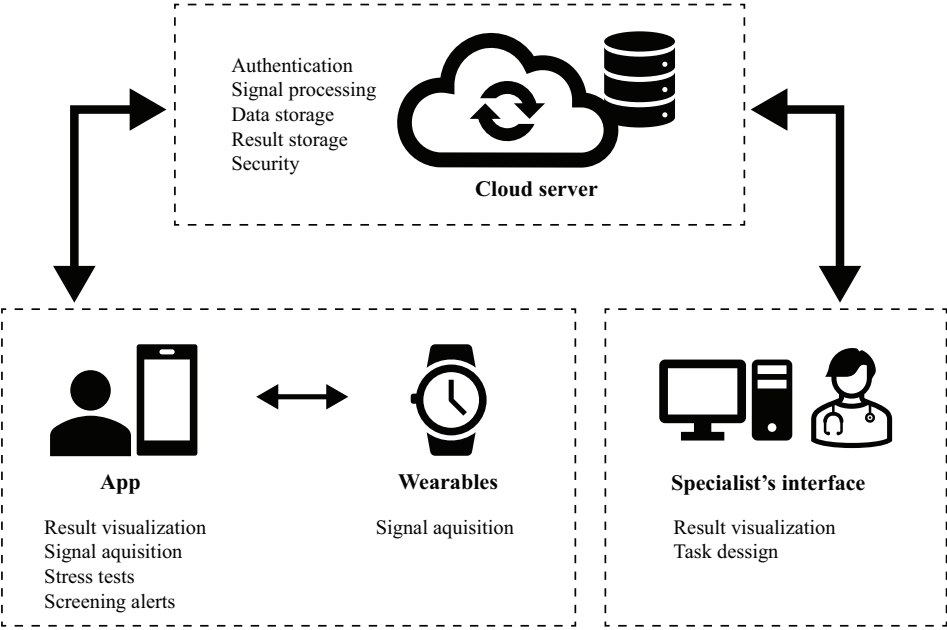


Figure 56: Service architecture. The cloud server authenticates users and specialists. Specialists are authorized to view solely the data of their respective patients.

Concept project (PDC2022-133197-I00). This service encompasses signal acquisition, processing, and user or specialist feedback. The architectural layout is illustrated in Figure 56.

The service encompasses a mobile application functioning as the user interface, enabling users to capture signals with SCPPG or Bluetooth-connected wearables, in addition to viewing their results or conducting stress tests to assess reactivity. Signals are partially processed on the smartphone, while most computation is performed in the cloud. The back-end handles user authentication, data storage, signal processing, and the storage and delivery of results.

Users can be either patients or nonpatients. For nonpatients, usage is limited to screening for sleep apnea with compatible wearable devices. For patients with depression or anxiety, they will be linked to their specialist. The specialist will be able to design protocols, stress tests, determine the



Figure 57: In-app stress tests. *Left:* Stroop test. The user’s voice is also recorded. *Right:* Trail making test. User’s lines are also recorded.

duration and frequency of the monitoring, and access the results. On the patient’s side, data acquisition includes signals during relaxation and stress tests (see Figure 57), surveys, diary entries, as well as the feedback selected by the specialist.

At this time, all devices of the Polar Electro brand (Kempele, Finland) are compatible. The Polar SDK has been used to establish a point-to-point connection between the app and the brand’s devices [1]. Figure 59 presents examples of three devices of this type. This approach enables the use of commercial wearable devices to acquire signals while processing is performed using the methods described in this thesis. Similarly to SCPPG, the signals can also be uploaded to the cloud for processing, thereby reducing the computational load on the smartphone. For example, the PPG signal obtained from any of the smartwatches, the OH1, or the Verity Sense can be utilized for OSA screening.

Over time, additional devices from other brands are expected to be inte-

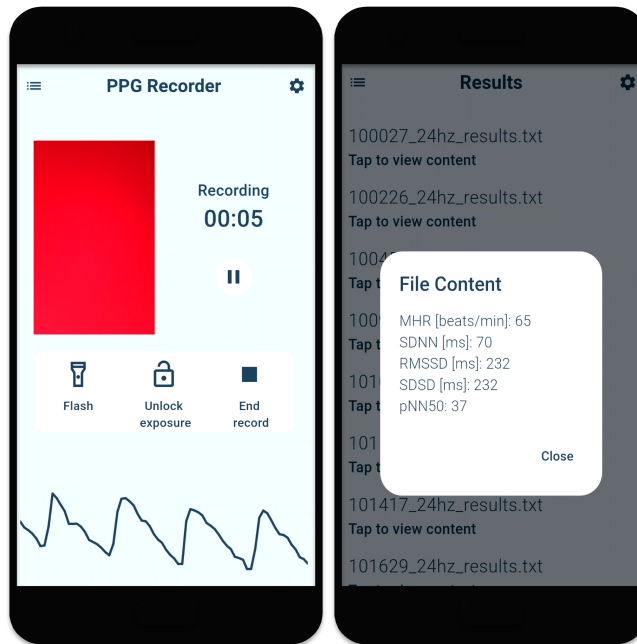


Figure 58: Smartphone PPG Recorder. *Left:* Smartphone PPG recorder. Camera feed, controllers and real-time signal. The signal is recorded and uploaded to cloud at the end of the session. *Right:* Visualization of PRV metrics after cloud computing. MHR, SDNN, RMSSD, SDSD and pNN50 are shown.

grated. This is a labor-intensive process, as each SDK must be incorporated individually and not all brands provide such services. An alternative approach is to listen to the Bluetooth Low Energy broadcast beacons that these devices typically emit. However, due to the limited bandwidth of this type of connection, only minimal information is transmitted. Typically, only the MHR calculated by the device is available, which means that the raw signals are not accessible for further processing to extract variability metrics.

Currently, two studies related to the app are being proposed in the context of depression and anxiety monitoring applications. The first study will be conducted in collaboration with the Human Openware Research Lab (HOWlab) at the Aragón Institute of Engineering Research (I3A) and the SOS Adolescents Association of Huesca. This study will involve designing and developing new dedicated services for patients and specialists

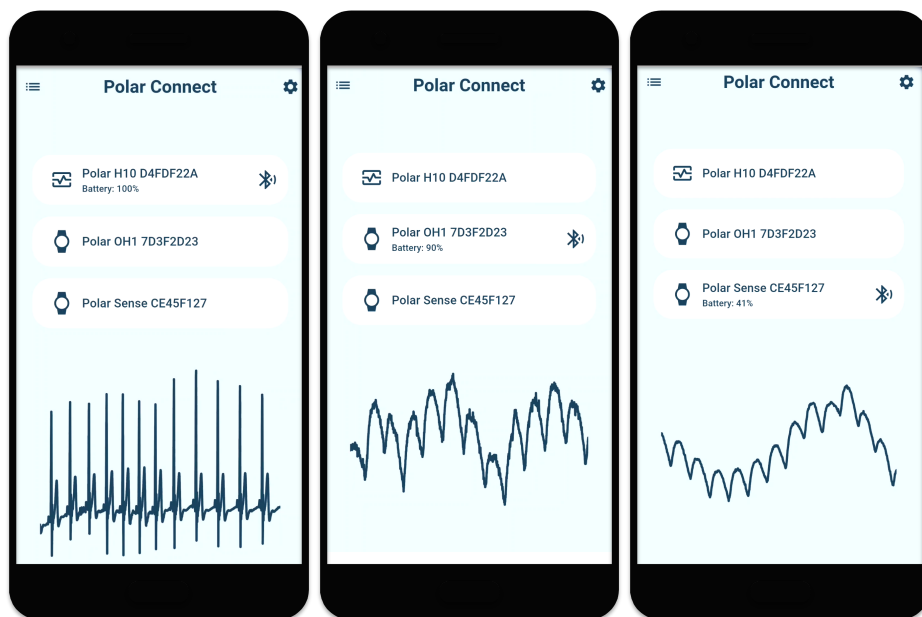


Figure 59: Polar Connect. Visualization of various signals from Polar devices. *Left:* ECG from Polar H10, *Center:* PPG from Polar OH1. *Right:* PPG from Polar Verity Sense.

to facilitate usability, adherence, and follow-up. Specialists from SOS Adolescentes will use the platform, integrating it into their regular workflow. The study will assess whether the tool is effective in monitoring and personalizing treatment, as well as evaluating any improvements suggested for future iterations.

The second study will be carried out in collaboration with the Mental Disorders Impact and Prevention Research Group at the Sant Joan de Déu Research Institute, affiliated with CIBERSAM. This study will investigate the most effective approach to assess stress reactivity in patient populations. Among other objectives, it will focus on determining the optimal frequency of stress-inducing tests to prevent patients from adapting to the stimulus while maintaining adequate sampling.

5 Conclusions

This thesis investigates a robust methodology for evaluating the autonomic nervous system using wearable devices. This evaluation was performed primarily using heart rate variability and pulse rate variability techniques, applied to research on chronic conditions such as obstructive sleep apnea, anxiety, and depression. In addition, a preliminary study of the peripheral venous pressure signal was performed to assess its potential for derived indices related to the parasympathetic system.

The research was implemented through the development of a smartphone app-based service capable of recording photoplethysmographic pulse signals using the camera, connecting to third-party wearable devices, conducting stress reactivity assessment protocols, and displaying results. The service also includes integration with a cloud server capable of user authentication, data storage, signal processing using the methods described in this thesis, and sending results back to the user.

The objectives outlined in the Introduction have been achieved as follows:

1. *General methods for reliable ANS assessment using wearable devices*

A methodology has been proposed for the robust evaluation of the autonomic nervous system using wearables, where each block in the processing chain is described as part of a coherent system. This processing chain requires a holistic perspective of all subprocesses to ensure a coordinated strategy. In this case, a restrictive approach to artifact removal was adopted to minimize false positives in detection. This ensures a reliable set of detections from which heart rate variability metrics can be derived.

A comparison of different techniques for calculating these metrics was conducted, analyzing their degradation in the presence of data loss caused by artifacts or poor signal-to-noise ratios. Based on the results, the most suitable methods have been proposed for each metric and level of data loss.

2. *Smartphone app development for PPG recording*

A smartphone application has been developed to obtain pulse photoplethysmography using the camera. This approach offers great versatility for conducting studies, as smartphones are universal devices. It also ensures good user acceptance, since it eliminates the need for additional device expenditures. Although the initial aim was not for this application to perform additional tasks, its potential led to an expansion of its functionalities, making it a central hub for the other research efforts included in this thesis.

The current system integrates signal processing with a cloud server and connects to third-party devices, enabling continuous monitoring.

3. *Application of wearables to mental illness*

The smartphone application has been used to monitor patients with depression and anxiety by assessing their stress reactivity. This evaluation is performed using the signal obtained from the smartphone camera, which is processed to derive heart rate variability metrics. The results reveal significant differences in stress reactivity compared to the control group. Specifically, patients with depression and anxiety exhibit reduced reactivity. Although this phenomenon has previously been studied, this is the first time it has been conducted using only a smartphone.

4. *Application of wearables to obstructive sleep apnea screening*

A novel index has been proposed that correlates with the apnea-hypopnea index, the standard for diagnosing and stratifying the severity of obstructive sleep apnea. This index offers a significant advantage because it is derived exclusively from wearable device signals, in contrast to the bulky sensors used in polysomnography. Moreover, due to its calculation method, it demonstrates a high degree of robustness compared to similar metrics.

Following a study using a large dataset obtained with non-wearable devices, a preliminary investigation was conducted with a wrist-worn device, yielding promising results despite additional challenges.

These challenges included the use of lower-quality signals (reflective photoplethysmography instead of transmissive, measurement at the wrist instead of the finger, increased motion artifacts, etc.) and the estimation of oxygen saturation from the same signal without prior calibration.

In conclusion, all the objectives proposed have been achieved. The findings of this thesis advance the methodology for evaluating the autonomic nervous system as a coherent and robust framework. Its applications are directly relevant to the healthcare field, as demonstrated by its use in obstructive sleep apnea and mental health conditions. Finally, the development of the smartphone app-based service bridges the gap between theory and practice.

Bibliography

- [1] Polar Electro. Polar SDK (github.com), 12 2024.

List of Figures

- 1 **Examples of wearable devices that use PPG sensors.** *From top left, clockwise:* Apple Watch PPG sensor using green LEDs; Polar OH1+ armband; a smartwatch sharing pulse rate estimations with a smartphone; and a runner using a fitness tracker. From [11–14]..... 15
- 2 **Operation of a reflection PPG.** This is the type of PPG provided by smartwatches, smartphones and devices placed on the forehead, arm and temple. It is opposed to the transmittance PPG, where the photodiode is facing the LED, provided by fingertip pulse oximeters and earlobe devices. Adapted from [15]..... 16
- 3 **ECG waveform.** The P wave corresponds with atrial depolarization; the QRS complex represents ventricular depolarization; and the T wave corresponds to ventricular repolarization. The R-R interval is the standard metric for measuring the time between heartbeats. 17
- 4 **Comparison of CVP and PVP and relation with the ECG.** CVP is composed of several waves: a (atrial contraction), c (tricuspid valve elevation into the right atrium), x (downward movement of the contracted right ventricle), v (back-pressure wave from blood filling the right atrium) and y (tricuspid valve opens in early ventricular diastole). These waves are not present in PVP, which has a pulsatile waveform similar to the PPG..... 20

-
- | | | | |
|----|-----------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 5 | NIVA device. | This device designed by Volumetrix (Nashville, TN) measure the changes in PVP in the wrist using a piezoelectric sensor. From [26]. | 21 |
| 6 | Sympathetic and Parasympathetic Pathways. | From [31]. | 24 |
| 7 | Illustration of the short-term ANS regulatory mechanisms of the cardiovascular system. | Vectorized and adapted from [39]. | 26 |
| 8 | Cardiac conduction and innervation. | From [32]. <i>Left:</i> Electrical conduction pathways. <i>Right:</i> Autonomic innervation. | 28 |
| 9 | Example of a tachogram. | The R-R interval physiologically fluctuates around a mean heart rate. | 30 |
| 10 | Airway obstruction during OSA. | Extracted from [33]. | 49 |
| 11 | Processing chain for HRV metrics. | Essential blocks in green, optional in grey. Within the framework of comprehensive ANS evaluation employing wearable technology, the components that are optional in controlled environments become essential. | 66 |
| 12 | Erroneous detection in an ECG affected by artifacts. | <i>Top:</i> Corrupted ECG. The artifact has been simulated in order to compare the actual beats with the automatically detected ones. <i>Bottom:</i> The filtered ECG is shown in black, while the clean ECG appears in gray. Green dots indicate correct detections, and red dots represent incorrect ones. The third heartbeat is missed due to a significant reduction in the QRS amplitude. The fifth heartbeat is misaligned because the artifact mimics a short-duration peak, masking the QRS complex. | 68 |

- 13 **Artifact identification using sliding windows of 10 seconds.** ($N = 2560$ at 256 Hz sampling rate). PPG in black, thresholds in red. In the event that any sample of the PPG within the 10-second interval surpasses the threshold, the entire interval is rejected (red shade). *First:* $\tilde{x}^{th}[n]$ threshold with $K_1 = 1.3$. *Second:* $x_\sigma^{th}[n]$ threshold with $K_2 = 1.5$. *Third:* $x_{MAD}^{th}[n]$ threshold with $K_3 = 2$. 72

- 14 **Artifact identification using Hjorth parameters.** *First:* PPG shown in black, with automatically identified artifacts highlighted in red. Since the Hjorth parameters are calculated over a window, some uncorrupted signal segments are also eliminated. *Second, Third and Fourth:* Dashed lines with dots represent the Hjorth parameters. Note that they are sampled at each dot, representing the window center, contrary to the signal. Unallowed regions are shaded with red. 74

- 15 **PPG pulse detection.** *Top:* PPG signal in black with event detections in red. *Bottom:* PPG derivative in black with event detections in red. The green line shows the adaptive threshold, which is divided into three distinct phases. Following each peak detection, the threshold stays steady for a refractory period. Next, it linearly declines to a minimum value, set as a fraction of the previous maximum. Lastly, the threshold remains fixed until the subsequent detection. 78

- 16 **Outlier detection in the event series.** Events located within the gray shaded area are incorrectly identified due to the artifact influence. *Top:* PPG signal. Detected pulses are marked with red dots. *Center:* PPG derivative. *Bottom:* Outlier detection in the interval function. Red shaded areas delimits values outside the allowed range. Intervals in red stems have been identified as outliers; however, some are still undetected. 81

- 17 **PSD estimates from the same record using different methods.** The estimates exhibit substantial differences. *Top:* Fourier-based spectral estimation. *Bottom:* Averaged Lomb-Scargle periodograms. 84
- 18 **IPFM model diagram.** The input function $m(t)$ modulates the variability, resulting in the event series t_k 85
- 19 **RSA band delimitation based in respiratory rate.** Bands are centered in the respiratory rate (red dot), estimated from respiratory signals or other signals with respiratory modulation. *Top:* Fixed width. In this example, the bandwidth is 0.1 Hz. *Bottom:* -3 dB bandwidth. 89
- 20 **RSA band delimitation based in spectral coherence.** *Top:* Spectral coherence between HRV and a respiration-modulated signal. The dashed line represents the threshold that defines the RSA band. *Center:* HRV spectral density. *Bottom:* Spectral density of a respiration-modulated signal. In this example, the slope transit time (see Chapter 5). 91
- 21 **OSP decomposition.** *Top panel:* Respiratory rate outside LF band. *Bottom panel:* Respiratory rate within LF band. *Inside each panel, left side top to bottom:* Respiratory signal, modulation of the heart rate, modulation linearly related with respiration, and residual component. *Inside each panel, right side top:* Respiratory signal (blue) and modulation of the heart rate (gray) spectra. *Inside each panel, right side bottom:* Modulation of the heart rate (gray), modulation linearly related with respiration (black), and residual component (dotted) spectra. 93
- 22 **Lagged Poincaré plots.** *Left:* Lagged Poincaré Plot with one-beat lag. *Right:* Ellipse fitting method. 96

- 23 **PPG and PVP delineation.** Key points include the maximum upslope ($t_U[k]$) marked by red circles, the pulse onset ($t_O[k]$) marked by green triangles, and the derivative zero crossing ($t_A[k]$) marked by blue triangles. The PPG signal is represented by a solid line, while its derivative is shown with a dashed line. *Bottom:* PVP delineation. 107
- 24 **Signed rank test results.** The test measures differences in median distributions of P_{HF} at different respiratory rates (6, 12 and 18 breaths per minute). Statistical significance is marked with an asterisk (*). *Top left:* Fixed width HF window, centered on respiratory rate. *Top right:* -3 dB width HF window, centered on respiratory rate. *Bottom:* Coherence-based HF window. ... 111
- 25 **ECG with missed beat detections due to noise and artifacts.** R-wave detections are marked with red circles. *Top:* Scattered missing beats due to a low SNR. *Bottom:* Burst of missing beats caused by an artifact. 122
- 26 **Example of simulation with a segment of 40 beats.** Deleted beats are displayed in red. *Top:* Random distributed missing beats, $p = 0.25$. *Bottom:* Bursts of missing beats. The elements at the ends (green) cannot be deleted. 123
- 27 **Process flow.** The first row describes common event-series processing steps, including outlier detection and rejection, as well as gap-filling methods. After these common steps, the different HRV metrics, grouped by families, are presented in each row. \mathcal{OD} = Outlier Detection; \mathcal{OR} = Outlier Rejection; \mathcal{L} = Linear; \mathcal{NL} = Non-Linear; \mathcal{M} = Model-based. 125
- 28 **Gap-filling algorithm flowchart.** Gray ovals represent terminal points, red rectangles represent processes, gray rectangles represent iterations, diamonds represent decisions, and green rectangles represent success in each gap correction. 127

- 29 **Demonstration of the \mathcal{L} gap-filling method.** The method prioritizes filling small gaps first. *Top*: Detected R-R intervals. The dashed line represents the moving median threshold, with intervals above it (in red) requiring modification. *Center*: End of the first iteration. Successfully corrected gaps are shown in green. An uncorrected gap (in orange) remains due to an insufficient number of insertions. This gap will be addressed in subsequent iterations. *Bottom*: End of the second iteration. The gap at sample 33 is now correctly filled with the appropriate number of insertions. 128
- 30 **Coverage of time-domain metrics from Apple Watch dataset.** The graphs show the percentage of cases (n_{th}) with a relative error below a certain threshold (ϵ_{th}). *Top left*: MHR. *Top right*: SDNN. *Bottom*: RMSSD. 135
- 31 **Relax (green)/Stress (blue) discrimination of time-domain metrics from Apple Watch dataset.** *Top left*: MHR. *Top right*: SDNN. *Bottom*: RMSSD. *: Significant differences ($p < 0.05$) between relax and stress groups. **: Significant differences ($p < 0.001$) between relax and stress groups. 136
- 32 **Coverage of frequency-domain metrics computed via Fourier transform from Apple Watch dataset.** The graphs show the percentage of cases (n_{th}) with a relative error below a certain threshold (ϵ_{th}). *Left*: P_{LF} . *Right*: P_{HF} 140
- 33 **Relax (green)/Stress (blue) discrimination of frequency-domain metrics computed via Fourier transform from Apple Watch dataset.** *Left*: P_{LF} . *Right*: P_{HF} . *: Significant differences ($p < 0.05$) between relax and stress groups. **: Significant differences ($p < 0.001$) between relax and stress groups. 140

- 34 **Coverage of frequency-domain metrics computed via Lomb-Scargle periodograms from Apple Watch dataset.** The graphs show the percentage of cases (n_{th}) with a relative error below a certain threshold (ϵ_{th}). *Left: P_{LF} . Right: P_{HF} .*..... 143
- 35 **Relax (green)/Stress (blue) discrimination of frequency-domain metrics computed via Lomb-Scargle periodograms from Apple Watch dataset.** *Left: P_{LF} . Right: P_{HF} . *: Significant differences ($p < 0.05$) between relax and stress groups. **: Significant differences ($p < 0.001$) between relax and stress groups.*..... 144
- 36 **Coverage of Poincaré metrics from Apple Watch dataset.** The graphs show the percentage of cases (n_{th}) with a relative error below a certain threshold (ϵ_{th}). *Top left: SD1. Top right: SD2. Bottom left: Md. Bottom right: Sd.*..... 147
- 37 **Relax (green)/Stress (blue) discrimination of Poincaré metrics from Apple Watch dataset.** *Top left: SD1. Top right: SD2. Center left: S. Center right: Md. Bottom: Sd. *: Significant differences ($p < 0.05$) between relax and stress groups. **: Significant differences ($p < 0.001$) between relax and stress groups.*..... 148
- 38 **Fourier transform spectral estimates from the same record with different loss rates of scattered missing beats.** *Top: \mathcal{M} correction. Bottom: \mathcal{L} gap filling.*..... 152
- 39 **Fourier transform spectral estimates from the same record with different burst durations.** *Top: \mathcal{M} correction. Bottom: \mathcal{L} gap filling.*..... 152
- 40 **Lomb-Scargle spectral estimates from the same record with different loss rates of scattered missing beats.** *Top: \mathcal{OR} correction. Bottom: \mathcal{L} gap filling.*..... 153

- 41 **Lomb-Scargle spectral estimates from the same record with different burst durations.** *Top: OR correction. Bottom: \mathcal{L} gap filling.* 153
- 42 SCPPG (black) and Reference (blue) signals. The image consists of 10-second segments extracted from different stages of the protocol, which have been assembled together. The time axis is shown as a reference; it does not indicate continuity..... 165
- 43 **SCPPG (black) and Reference (blue) modulating signals $m(t)$ obtained with the TVIPFM model for each stage.** The agreement is very high at all stages, especially at rest..... 169
- 44 **Regression and Bland-Altman plots of relative differences for each metric.** All stages together. *First: MHR. Second: SDNN. Third: RMSSD. Fourth: $P_{\perp,LF}$. Fifth: $P_{r,LF+HF}$.* 173
- 44 **Boxplots for each metric. Significant differences are marked with asterisks: * for $p < 0.05$ and ** for $p < 0.001$. Values after asterisks indicate the effect size.** *First: MHR. Second: SDNN. Third: RMSSD. Fourth: \mathcal{R}' .* 175
- 45 **Signal processing flowchart of signals employed in segment classification (left) and annotation (right).** Gray boxes represent signals and blue boxes represent processes. 194
- 46 *Top panel: Event-based annotations.* Solid lines represent nasal pressure (top row) and airflow (bottom row). The dashed line represents basal respiration. The onsets and endsets of the annotation correspond to the reductions in airflow and the restorations, respectively. *Bottom panel: Grouped annotations.* Events are grouped in apneic/hypopneic bursts. Note that new bursts are initiated every 8 events, allowing a precise apneic/hypopneic characterization. These annotations are then transformed into segment-based annotations..... 196

- 47 **Segment classification models.** Gray boxes represent signals, blue boxes represent processes and red boxes represent models. 198
- 48 **AHI v. CVHRI after binary segment classification.** Vertical lines separate $AHI < 5$, $5 \leq AHI < 15$ and $AHI \geq 15$ groups. *Top left:* Segment classification using PPI + SpO_2 . *Top right:* Segment classification using SpO_2 . *Bottom:* Segment classification using PPI. 205
- 49 **AHI v. CVHRI after multiclass segment classification.** Vertical lines separate $AHI < 5$, $5 \leq AHI < 15$ and $AHI \geq 15$ groups. *Top left:* Segment classification using PPI + SpO_2 . *Top right:* Segment classification using SpO_2 . *Bottom:* Segment classification using PPI. 206
- 50 **Maxim MAXREFDES103 [1].** A wrist-worn device capable of recording PPG with red, green, and infrared channels. 222
- 51 **Upper (blue) and lower (green) envelopes of the PPG.** Blue triangles indicate $t_O[k]$ points, while green triangles indicate $t_A[k]$ points. 223
- 52 **Outlier removal in R** *Top:* R signal in black, outlier shaded in red. *Bottom:* Squared derivative of R in solid black. Dashed line corresponds to the fixed threshold at 10^{-8} 224
- 53 **SpO_2 estimation from Maxim.** *Top:* Normalized R from Maxim. *Bottom:* SpO_2 from reference in green, estimate from Maxim in black. 225
- 54 **AHI v. CVHRI from a wearable device after binary segment classification.** Vertical lines separate $AHI < 5$, $5 \leq AHI < 15$ and $AHI \geq 15$ groups. 226
- 55 **AHI v. CVHRI from a wearable device after binary segment classification.** Vertical lines separate $AHI < 5$, $5 \leq AHI < 15$ and $AHI \geq 15$ groups. 227

56 **Service architecture.** The cloud server authenticates users and specialists. Specialists are authorized to view solely the data of their respective patients. 238

57 **In-app stress tests.** *Left:* Stroop test. The user’s voice is also recorded. *Right:* Trail making test. User’s lines are also recorded.239

58 **Smartphone PPG Recorder.** *Left:* Smartphone PPG recorder. Camera feed, controllers and real-time signal. The signal is recorded and uploaded to cloud at the end of the session. *Right:* Visualization of PRV metrics after cloud computing. MHR, SDNN, RMSSD, SDSD and pNN50 are shown. 240

59 **Polar Connect.** Visualization of various signals from Polar devices. *Left:* ECG from Polar H10, *Center:* PPG from Polar OH1. *Right:* PPG from Polar Verity Sense. 241

List of Tables

- 1 **Autonomic effects on various organs** [32]. Sympathetic and parasympathetic branches usually have opposite effects on the same tissue..... 23
- 2 **Respiratory rate estimation from PVP and PPG**..... 110
- 3 **Relative error (%) of time-domain metrics.** (a) Scattered missing beats. (b) Bursts. †: Significant differences ($p < 0.05$) between \mathcal{OR} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{OR} 132
- 4 **p -values of ranked signed test for supine/tilt discrimination of time-domain metrics.** N.S.: Not Significant ($p > 0.05$).... 133
- 5 **Relative error (%) of time-domain metrics from Apple Watch dataset.** †: Significant differences ($p < 0.05$) between \mathcal{OR} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{OR} 134
- 6 **Relative error (%) of frequency-domain metrics computed via FFT.** (a) Scattered missing beats. (b) Bursts. †: Significant differences ($p < 0.05$) between \mathcal{M} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{M} 137

- 7 **p -values of ranked signed test for supine/tilt discrimination of frequency-domain metrics computed via FFT.** N.S.: Not Significant ($p > 0.05$). 138
- 8 **Relative error (%) of frequency-domain metrics computed via FFT from Apple Watch dataset.** †: Significant differences ($p < 0.05$) between \mathcal{M} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{M} 139
- 9 **Relative error (%) of frequency-domain metrics computed via Lomb-Scargle periodograms.** (a) Scattered missing beats. (b) Bursts. †: Significant differences ($p < 0.05$) between \mathcal{OR} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{OR} 141
- 10 **p -values of ranked signed test for supine/tilt discrimination of frequency-domain metrics computed via Lomb's method.** N.S.: Not Significant ($p > 0.05$). 142
- 11 **Relative error (%) of frequency-domain metrics computed via Lomb-Scargle periodograms metrics from Apple Watch dataset.** †: Significant differences ($p < 0.05$) between \mathcal{OR} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{OR} 143
- 12 **Relative error (%) of Poincaré metrics.** (a) Scattered missing beats. (b) Bursts. †: Significant differences ($p < 0.05$) between \mathcal{OR} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{OR} . 145
- 13 **p -values of ranked signed test for supine/tilt discrimination of Poincaré metrics.** N.S.: Not Significant ($p > 0.05$). 146

14	Relative error (%) of Poincaré metrics from Apple Watch dataset. †: Significant differences ($p < 0.05$) between \mathcal{OR} and \mathcal{L} . Δ : Significant differences ($p < 0.05$) between \mathcal{L} and \mathcal{NL} . §: Significant differences ($p < 0.05$) between \mathcal{NL} and \mathcal{OR}	147
15	Summary of findings. Best correction method and maximum acceptable missing beats for a relative error less than 20% in the third quartile.	149
16	Demographic characteristics. Significant differences between paired controls and patients ($p < 0.05$) in bold. <i>SD</i> = Standard Deviation; <i>N</i> = Number; BMI = Body Mass Index; PSS = Perceived Stress Scale; STAI = State-Trait Anxiety Inventory. . .	167
17	Percentage of artifact-free time according to Hjorth-based automatic detection.	171
18	Agreement results of PRV metrics. Median and standard deviation, error median and standard deviation, Pearson correlation coefficient, number of missing cases, and Bland-Altman mean difference. All subjects.	172
19	Binary segment classification metrics.	201
20	Multiclass segment classification metrics.	202
21	Binary segment classification results (%)	204
22	Multiclass segment classification results (%)	204
23	Stratification after binary segment classification (% , except AUC and κ).	205
24	Stratification after multiclass segment classification (% , except AUC and κ).	206

25	Multiclass segment classification results (%) . Comparison between using/omitting <i>severe hypopnea</i> label using the PPI+SpO ₂ model.	208
26	Comparison with other studies (% , except AUC and κ).	210
27	Binary segment classification results including frequency-domain PRV metrics as predictors (%)	214
28	Multiclass segment classification results including frequency-domain PRV metrics as predictors (%)	214
29	Binary segment classification results from a wearable device (%)	226
30	Binary segment classification results using PPI from wearable device and SpO₂ from reference (%)	226

List of Acronyms

AASM	American Academy of Sleep Medicine
ANS	Autonomic Nervous System
AUC	Area Under the Curve
AV	Atrioventricular (node)
BPM	Beats Per Minute
CNN	Convolutional Neural Network
CNS	Central Nervous System
CVHR	Cyclic Variation of the Heart Rate
CVHRI	Cyclic Variation of the Heart Rate Index
DAP	Decreases of Amplitude fluctuations of the PPG
ECG	Electrocardiography
GAD	Generalized Anxiety Disorder
HF	High Frequency
HRV	Heart Rate Variability
IBI	Inter-Beat Interval
IPFM	Integral Pulse Frequency Modulation (model)
L	Linear (correction method)
LF	Low Frequency
LSTM	Long Short-Term Memory (model)

M	Model-based (correction method)
MDD	Major Depressive Disorder
MHR	Mean Heart Rate
NL	Non-Linear (correction method)
OD	Outlier Detection
OR	Outlier Removal
OSA	Obstructive Sleep Apnea
PPG	Pulse Photoplethysmography
PPI	Pulse-to-Pulse Interval
PRV	Pulse Rate Variability
RMSSD	Root Mean Square of Successive Differences
ROC	Receiver Operating Characteristics (curve)
RSA	Respiratory Sinus Arrhythmia
SA	Sinoatrial (node)
SCPPG	Smartphone Camera Pulse Photoplethysmography
SDNN	Standard Deviation of Normal-to-Normal intervals
SNR	Signal-to-Noise Ratio
TVIPFM	Time-Varying Integral Pulse Frequency Modulation (model)
VLF	Very Low Frequency

List of Publications

The research carried out during the Ph.D. led to the production of the following publications.

Journal Articles

- **Diego Cajal**, David Hernando, Jesús Lázaró, Pablo Laguna, Eduardo Gil & Raquel Bailón. *Effects of Missing Data on Heart Rate Variability Metrics*. Sensors 2022, vol. 22, no. 15, 5774. **Corresponding author.**
- **Diego Cajal**, Eduardo Gil, Pablo Laguna, Carolina Varon, Dries Testelmans, Bertien Buyse, Chris Jensen, Rohan Hoare, Raquel Bailón & Jesús Lázaró. *Obstructive Sleep Apnea Screening by Joint Saturation Signal Analysis and PPG-derived Pulse Rate Oscillations*. IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 1, pp. 228-238, 2024. **Journal article. Corresponding author.**
- **Diego Cajal**, Concepción de la Cámara, Mar Posadas-de Miguel, Noel Torrijos, Óscar Nadal, Teresa Blanco, Sara Siddi, Pablo Armañac, Eduardo Gil, Jesús Lázaró & Raquel Bailón. *Evaluation of Stress Response Using Smartphone PPG for Anxiety and Depression Monitoring*. **Corresponding author. Pending revision.**

International Conferences

- **Diego Cajal**, David Hernando, Jesús Lázaro, Eduardo Gil, Annie Alvis, Monica Polcz, Bret Alvis, Kyle Hocking, Colleen Brophy & Raquel Bailón. *Parasympathetic Characterization Guided by Respiration From Wrist Peripheral Venous Pressure Waveform*. 2020 Computing in Cardiology, Rimini, Italy, 2020, pp. 1-4. **Oral communication.**
- **Diego Cajal**, Mar Posadas-de Miguel, Concepción de la Cámara, Spyridon Kontaxis, Jesús Lázaro & Raquel Bailón. *Smartphone PPG Validation for a Depression Assessment Protocol*. 2022 IEEE e-Health and Bioengineering Conference (EHB), Iasi, Romania, 2022, pp. 1-4. **Oral communication.**
- **Diego Cajal**, Eduardo Gil, Pablo Laguna, Carolina Varon, Dries Testelmans, Bertien Buyse, Chris Jensen, Rohan Hoare, Raquel Bailón & Jesús Lázaro. *Sleep apnea severity stratification by an FFT-based PPG-derived index*. 2022 12th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO), Vysoké Tatry, Štrbské Pleso, Slovakia, 2022, pp. 1-2. **Oral communication.**

National Conferences

- **Diego Cajal**, Eduardo Gil, Pablo Laguna, Carolina Varon, Dries Testelmans, Bertien Buyse, Chris Jensen, Rohan Hoare, Raquel Bailón & Jesús Lázaro. *Estratificación de la severidad de la apnea del sueño mediante un índice derivado de PPG basado en FFT*. Jornada de Jóvenes Investigadores del I3A, vol. 10 (Actas de la XI Jornada de Jóvenes Investigadores del I3A – 16 de junio de 2022) **Oral communication.**
- **Diego Cajal**, Mar Posadas-de Miguel, Concepción de la Cámara, Spyridon Kontaxis, Jesús Lázaro & Raquel Bailón. *Smartphone PPG Validation for a Depression Assessment Protocol*. I Young Researchers Meeting CIBERESP-BBN 21-22 November 2022. **Poster.**