FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Analysis of the association of cardiac and respiratory signals

Ana Filipa da Costa Almeida

DISSERTATION

Integrated Master in Bioengineering

Supervisor: Teresa Sarmento Henriques

July 26, 2021

Analysis of the association of cardiac and respiratory signals

Ana Filipa da Costa Almeida

Integrated Master in Bioengineering

July 26, 2021

Resumo

Os sistemas fisiológicos podem ser caraterizados como caóticos e fractais sendo de natureza complexa e instável. Medidas de caos, fractalidade e complexidade estão normalmente associadas a sinais fisiológicos tais como a frequência cardíaca e a frequência respiratória.

A análise de formulações matemáticas que combinam estes conceitos é extremamente importante para compreender a informação contida nos sinais fisiológicos e desenvolver métodos para identificar possíveis problemas de saúde.

A sépsis neonatal é uma doença causada por uma resposta corporal desregulada a infeções que afeta os bebés especialmente no primeiro mês de vida e os recém-nascidos de baixo peso. Mundialmente, a incidência da sépsis neonatal pode variar entre 1 e 50 por 1000 bebés.

O diagnóstico precoce, tratamento e prevenção são cruciais, mas a sépsis neonatal é um enigma para os especialistas na área devido a mudanças na epidemiologia, à falta de marcadores de diagnóstico e à ausência de sinais evidentes. As alterações na frequência cardíaca e respiratória podem ser detetáveis na fase inicial da sépsis e, portanto, estes sinais podem ajudar a prever a sépsis neonatal.

Neste trabalho, foram aplicados métodos de processamento de sinal lineares e não lineares como a entropia e a compressão para analisar intervalos RR e sinais respiratórios para tentar prever a sépsis neonatal. Métodos como a *cross-entropy*, a *mutual information* e a *normalized compression distance* foram também utilizados para medir o acoplamento cardio-respiratório. Foi utilizado um conjunto de dados de bebés com e sem sepsis de um estudo internacional. Os sinais foram selecionados de três momentos diferentes: ao nascimento, perto da sépsis e após a sépsis. A ideia era comparar os dois grupos de bebés.

Não foram encontradas diferenças significativas para os três momentos. Apesar de isso ter sido esperado à nascença, algumas diferenças eram esperadas perto do momento de infeção. Os resultados sugerem que os métodos lineares têm um melhor desempenho na deteção de sépsis em oposição aos métodos não lineares, demonstrado pelo elevado nível de significância e tamanho do efeito obtidos para o RMSSD em particular.

Uma vez que o tamanho da amostra era pequeno, mais estudos deveriam ser realizados, com um conjunto de dados maior e métodos de processamento de sinais diferentes.

Abstract

Healthy living systems can be characterized as chaotic and fractal. Measures of chaos, fractality, and complexity are generally associated with physiological signals such as the heart rate and the respiratory rate.

The analysis of mathematical formulations that combine these concepts is fundamental to understand the information contained in physiological signals and to develop methods to identify possible health problems.

Neonatal sepsis is a life-threatening organ disease caused by a dysregulated body response to infection that affects infants, especially in the first month of life and low-birth-weight newborns. Worldwide, the incidence of neonatal sepsis can go from 1 to 50 per 1000 live births.

Early diagnosis, treatment, and prevention are crucial. However, neonatal sepsis is an enigma for specialists in the area due to changes in epidemiology, the lack of diagnostic markers, and the absence of obvious signs. Changes in heart rate and respiratory rate can be detectable in the early phase of sepsis, and, therefore, these signals can assist the prediction of neonatal sepsis.

In this work, linear and non-linear signal processing methods such as entropy and compression were applied to analyze RR intervals and respiratory signals to attempt to predict neonatal sepsis. Methods such as cross-entropy, mutual information, and normalized compression distance were also employed to measure cardio-respiratory coupling. A data set from neonates with and without sepsis from an international study was used. The signals were selected from three different moments: at birth, near sepsis and after sepsis. The idea was to compare the two groups of infants.

No significant differences were found for the three moments. Although this was expected at birth, some differences were expected near the moment of infection. The results suggest that linear methods have a better performance in the detection of sepsis in opposition to non-linear methods, shown by the high significance and effect size obtained for RMSSD in particular.

More studies should be performed since the sample size was small, with a larger data set and different signal processing methods.

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Teresa Henriques, for giving me the chance to work in this interesting project. This thesis wouldn't be possible without your help and guidance and it was a pleasure to meet you.

To all my teachers at FEUP and ICBAS, a big thank you for sharing your knowledge with me and providing me with the tools I needed to choose my path.

To my family, especially my mother and my sister, Sofia, I would like to thank you for always supporting and believing in me, even when I didn't really believe in myself. I wouldn't be who I am today if you weren't in my life. Also, a thank you to the rest of my family who have always given me nice words and congratulated me for my achievements.

To the friends I made before university even started, I would like to thank you for sticking with me and always encouraging me to do my best. Also, this experience wouldn't be the same without the friends I made along the way. I will never forget all the memories, the gatherings, the study sessions and the talks that made these five years a lot more enjoyable. A special thank you to Francisco, Diva and Leonor, who I have been friends with since first grade, Moutinho, Helena and Emanuel, who were an important part of my university experience, and Inês and Joana, who were my companions in the branch of Biomedical Engineering. Lastly but not less important, I would like to thank Ana and Rita, for accompanying me throughout these five years and always having my back. I couldn't have asked for better people in my life.

I am extremely grateful for everything and waiting for the next challenge.

Ana Filipa da Costa Almeida

"This is how you do it: you sit down at a keyboard and you put one word after another until it's done. It's that easy and that hard."

Neil Gaiman

Contents

1 Introduction **1**
- 1.1 Context and motivation . . . 1
- 1.2 Goals . . . 3
- 1.3 Outline . . . 3

2 Cardio-respiratory signals **5**
- 2.1 Heart rate . . . 5
- 2.2 Respiratory signal . . . 6
- 2.3 Cardio-respiratory coupling . . . 7
  - 2.3.1 Respiratory Sinus Arrhythmia . . . 7
  - 2.3.2 Phase synchronization . . . 8
- 2.4 Previous work . . . 9

3 Signal processing methods **11**
- 3.1 Time-domain analysis . . . 11
- 3.2 Non-linear analysis . . . 12
  - 3.2.1 Entropy-based methods . . . 12
  - 3.2.2 Mutual information . . . 14
  - 3.2.3 Normalized compression distance . . . 14
- 3.3 Methodologies . . . 16
- 3.4 Statistical analysis . . . 16

4 Time series **17**
- 4.1 Data information . . . 17
- 4.2 Signals' representation and resampling . . . 19

5 Results **23**
- 5.1 At birth . . . 23
- 5.2 Near sepsis . . . 25
- 5.3 After sepsis . . . 26

6 Discussion **29**

7 Conclusion **31**

References **33**

A Results: Correlation between compressors **37**

List of Figures

2.1 RR interval from an ECG. 6
2.2 Respiratory signal with a ttot interval represented. 6
2.3 Coordination between the heart rate and the respiratory signal. 8

4.1 Weeks for control babies signal collecting. 18
4.2 Weeks for sepsis babies signal collecting. 18
4.3 Example of the recorded signals. 19
4.4 Variation of MSE with different resampling frequencies. 20
4.5 Original and resampled RR signals. 20
4.6 Original and resampled RESP signals. 21

5.1 Example of the MSE obtained for a baby at birth. 24
5.2 Example of the MSC obtained for a baby at birth. 24

A.1 Correlation between the MSC_{scale1} of the three compressors. 37
A.2 Correlation between the NCD of the three compressors. 37

List of Tables

1.1 Risk factors of neonatal sepsis. 2

3.1 Interpretation of Cliff's Delta values. 16

4.1 Control data set. 17
4.2 Sepsis data set. 17
4.3 Average $\pm$ STD MSE of original and resampled signals. 20

5.1 Linear measures obtained for the RR and RESP signals at birth. 23
5.2 MSE measures obtained for the RR and RESP signals at birth. 24
5.3 MSC measures obtained for the RR and RESP signals at birth. 25
5.4 Joint non-linear measures obtained for the RR and RESP signals at birth. 25
5.5 Linear measures obtained for the RR and RESP signals near sepsis. 25
5.6 Non-linear measures obtained for the RR and RESP signals near sepsis. 26
5.7 Differences between the linear measures obtained at birth and near sepsis. 26
5.8 Differences between the non-linear measures obtained at birth and near sepsis. . . 27
5.9 Differences between the linear measures obtained near and after sepsis. 27
5.10 Differences between the non-linear measures obtained near and after sepsis. . . . 28

Abbreviations and Symbols

AMI	Auto mutual information
AMIF	Auto mutual information function
ANS	Autonomic nervous system
ApEn	Approximate entropy
CE	Cross-entropy
CMI	Cross mutual information
CMIF	Cross mutual information function
CRC	Cardio-respiratory coupling
ECG	Electrocardiogram
EEG	Electroencephalogram
EOS	Early-onset sepsis
HR	Heart rate
HRV	Heart-rate variability
ID	Information distance
LOS	Late-onset sepsis
MI	Mutual information
MSE	Multiscale entropy
MSC	Multiscale compression
NCD	Normalized compression distance
NID	Normalized information distance
PS	Phase synchronization
RESP	Respiratory
RSA	Respiratory sinus arrhythmia
SampEn	Sample entropy
STD	Standard deviation
ttot	Total respiratory time

Chapter 1

Introduction

1.1 Context and motivation

Healthy living systems can be characterized as chaotic and fractal in nature [1]. Human physiological signals are complex and not stable but fluctuate in time. The pattern of these fluctuations results from an attempt to maintain the system's equilibrium and can be treated by stochastic and chaotic non-linear mathematical models [2]. It is known that chaotic systems are non-linear and tend to produce what looks like random outputs. Some other characteristics include irregular behavior and abrupt transitions [3]. Nevertheless, they have a well-defined underlying order. Fractality shows the extent of the organization underlying chaotic structures describing its geometry. Therefore, both concepts of fractality and chaos are applied together [1].

Applying these non-linear analysis methods to physiological signals might be the solution to detect something wrong in the human body. Usually, physiological oscillators such as the electroencephalogram (EEG), heart rate, arterial pressure, and respiratory rate are used. The signals are recorded, becoming data that can then be stored and processed. Signal processing methods can focus on specific diseases, such as epilepsy, cardiovascular diseases, and diabetes [3]. In fact, studies show that fractality is associated with the state of health of a living organism [1].

Sepsis is a global healthcare issue that can be a leading cause of critical illness and hospital mortality. The definition of sepsis has changed around time. At this moment, sepsis is considered a life-threatening organ dysfunction caused by a dysregulated patient response to infection [4, 5]. The body's response causes injuries in its tissues and organs. Sepsis is known to modify non-immunologic pathways such as neuronal, metabolic, hormonal, autonomic and cardiovascular. Patients who survive sepsis might suffer from long-term sequelae with significant implications. Sepsis is the primary cause of death from infection due to the non-homeostatic body response. It might lead to septic shock when there is persisting hypotension associated [6]. Septic shock is considered a subset of sepsis where acute circulatory failure occurs and there is a much higher probability of death. Therefore, there is an importance in its early recognition and diagnosis.

Neonatal sepsis is a cause of substantial morbidity and mortality in infants, and it is known as a serious bacterial, viral, or fungal condition associated with various clinical manifestations [7].

Neonatal sepsis can be early-onset (EOS) when it appears within the first 72h after delivery or late-onset (LOS) when it appears from 3 to 7 days of age. The type of sepsis will help decide the best treatment since they come from different risk factors and pathogens. EOS usually results from the transmission of bacteria from mothers to newborns, while LOS comes from a horizontal transmission of infectious agents post-delivery [7, 8]. Table 1.1 shows the risk factors of both types of neonatal sepsis.

Table 1.1: Risk factors of neonatal sepsis [8].

	Early-onset sepsis	**Late-onset sepsis**
Risk factors	• Maternal infection • Premature or prolonged rupture of membranes • Multiple gestation • Prematurity	• Antibiotics • Low birth weight • Invasive procedures • Ventilator or catheter use • Prematurity

During pregnancy, the fetus is protected from pathogens in the amniotic sac, but this ends, giving them vulnerability during the birth. At birth, babies have not developed acquired immunity since they have not been exposed to pathogens, so they need to use the innate immune system to fight infection. The immaturity of the immunological system might lead to a reduced response to pathogens.

Premature newborns do not have a mature immune system, making it harder for them to defend themselves against infectious agents [9]. Therefore, premature and low birth-weight babies are at a higher risk of developing sepsis, more specifically, LOS, because they require prolonged hospitalization using different medical supplies [8].

Signs and symptoms of neonatal sepsis are not specific for this illness. They may include, among others, temperature instability leading to hypothermia or hyperthermia, abnormal heart rate (tachycardia or bradycardia), respiratory instability, apnea, irritability, continuously crying, seizures, abdominal distention, or bleeding [7, 9].

Worldwide, the incidence of neonatal sepsis can go from 1 to 50 per 1000 live births [10]. A study [11] from the US showed that from 1988 to 2006, approximately 2.5 million babies with less than 3 months were diagnosed with sepsis and, in 2006, the incidence was of 30.8 per 1000 live births. A study [10] found a sepsis incidence rate of 4 per 1000 live births between 2001 and 2006 in Taiwan and an incidence of 1.43 per 1000 live births between 2011 and 2015 in Switzerland. A more recent study [9] from the UK in 2016 found a rate of 8 per 1000 infants with neonatal sepsis.

Globally, it is estimated that there are 3.1 million newborn deaths annually and almost 1 million of those deaths are due to invasive infections such as neonatal sepsis [12]. Neonatal sepsis can cause death in 2% of term infants, 20% of premature newborns and 30% of babies diagnosed with meningitis [13]. A different study [12] presented a mortality rate of 51% for EOS and 45.9% for LOS from all infected babies.

As said, early diagnosis, treatment and prevention are crucial. However, neonatal sepsis is an enigma for specialists in the area due to changes in epidemiology, the lack of diagnostic markers, and the absence of apparent signs [8]. Since the initial signs are subtle and not specific, neonatal sepsis can be missed [13].

Isolation of bacteria from blood is considered the gold standard. However, the results can take up to 48h [8] or even 72h [14], according to different sources, and some routine laboratory tests have low sensitivity. In most neonatal sepsis cases, the diagnosis is based on a combination of clinical signs, risk factors and laboratory tests [11].

It is hypothesized that alterations in heart rate and respiratory rate are often detectable in the early phase of sepsis [15]. Thus, there is a need to find new ways to detect or even diagnose health issues, such as neonatal sepsis. Linear measures and non-linear methods such as cross-entropy (CE), mutual information (MI), and normalized compression distance (NCD) can be applied to measure the association between cardiac (RR) and respiratory (RESP) signals and detect neonatal sepsis.

1.2 Goals

This dissertation project aims to explore the interaction between the cardiac and the respiratory signals in the prediction of neonatal sepsis, creating a model to detect it. This will be achieved by applying different signal processing methods to the signals and comparing them to see which gives better results in the detection of infection.

Several tasks were needed to reach this goal: RESP and RR signals from infants at birth, before and after sepsis were collected and resampled. After that, signal processing methods were employed, and the linear and non-linear measures were applied to the signals. Statistical measures were used to evaluate the results and the measures were compared.

To conclude, this work studies approaches to attempt to detect sepsis in neonates.

1.3 Outline

This document is divided into seven chapters (including this chapter).

Chapter 2 describes the dynamics of the cardiac and respiratory signals and their coupling, as well as methods applied to them, found in the literature.

Chapter 3 explains the theory behind the methods that will be applied to the physiological signals during this work.

Chapter 4 describes the data set, the signals and how they were processed to be applied to the signal processing measures.

Chapter 5 shows and compares the results of the methods described previously.

Finally, chapter 6 presents a discussion of the results and chapter 7 the conclusion.

Chapter 2

Cardio-respiratory signals

This thesis is focused on exploring the dynamics of the cardiac and respiratory signals to help in the prediction of neonatal sepsis. This chapter is dedicated to explaining the signals' underlying concepts and how they interact in the human body.

2.1 Heart rate

The autonomic nervous system (ANS) is essential to control different physiological systems and is divided into sympathetic, parasympathetic, and enteral. The first two divisions are opposed, being the first accountable for stress reactions and the second for relaxing. Thus, for the heart, the sympathetic system is responsible for increasing heart frequency and contractility triggered by stress. The parasympathetic system maintains heart frequencies and contractility [16].

Heart rate (HR) is originated by the activity of the pacemaker cells located in the sino-atrial node in the right atrium of the heart and the rhythm is created by depolarization, and action potential [17].

Heart-rate variability (HRV) measures the difference between consecutive beat intervals, called the RR intervals, from the electrocardiogram (ECG) signal. It reflects the beat-to-beat changes in the heart's rhythm in response to stimuli [15, 18]. Figure 2.1 shows the definition of a RR interval. High HR is associated with lower variability in RR intervals [19]. Besides that, breathing and HR coordination are related to HRV [3].

Analyzing the HRV has been proved to be a valuable tool in biomedical and clinical research as it can be used as a marker for cardiovascular health. A healthy cardiovascular system is associated with a chaotic HRV with non-linear properties. A decreased HRV can be an indication of heart failure [21].

HRV analysis can also be used for the diagnosis of neonatal diseases, although it is challenging to record long-term signals in premature infants due to their unstable conditions and unpredictable movements. HRV can show important information about premature infants' status helping in decision support systems [19, 18].

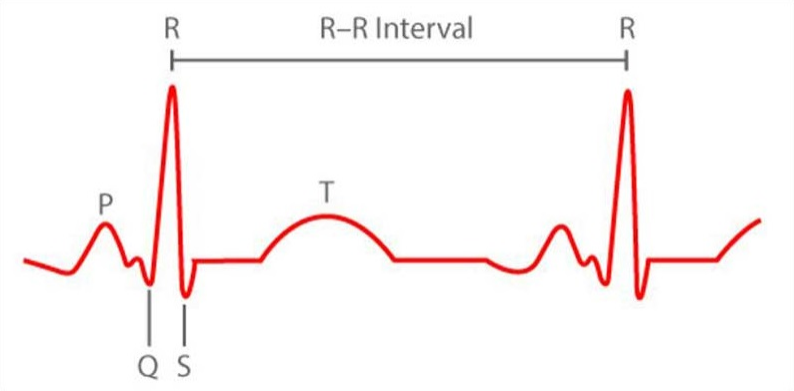


Figure 2.1: RR interval from an ECG [20].

2.2 Respiratory signal

The respiratory system's primary function is to maintain homeostasis between the exterior and interior of the body, where the ANS plays an important role [22].

Unlike the heart rate, the respiratory rhythm is not originated from the pacemaker cells but within the brainstem. It consists of multiple neurons with a characteristic discharge pattern through the two phases of respiration (inspiration and expiration) [17]. During inspiration, negative intrathoracic pressure is generated, and, during expiration, the pressure in the thorax is equaled to the exterior pressure [22].

It is known that different sleep states create different effects on respiration. For the case of infants, the respiratory rate is lower at sleep in comparison to the wake state and more variable at active sleep in comparison to quiet sleep [23].

Additionally, periodic breathing is common in newborns and respiratory rate decreases during the first 6 months of life. In some cases, it may be necessary an increased respiratory rate to maintain ventilation with minimal effort.

Figure 2.2 represents a total respiratory time (ttot) interval, which is the difference between two consecutive minimums in a respiratory signal. The breathing cycle duration, used in this work by the name of RESP signal, corresponds to the total ttot time series.

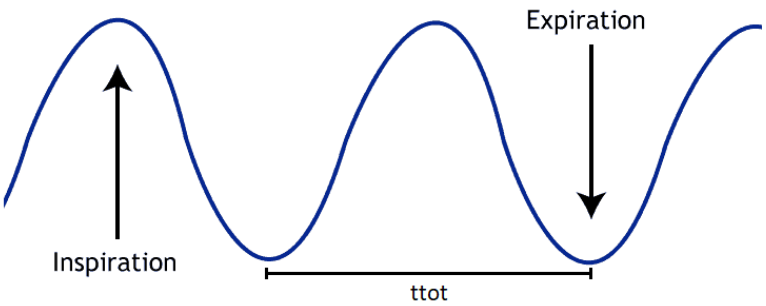


Figure 2.2: Respiratory signal with a ttot interval represented.

2.3 Cardio-respiratory coupling

It is known that the cardiovascular and respiratory systems are not independent [24]. These interactions are important for the regulation of blood gases and central nervous functions [21].

Due to the complexity of both cardiac and respiratory signals having transient and non-linear characteristics associated with different states and conditions, it is complicated to evaluate their interactions. Cardio-respiratory coupling (CRC) can be obtained by the analysis of both physiological signals, but its relevance is not well known [25, 26].

CRC has multiple levels. Besides originating the respiratory rhythm, the brainstem is also provided with information regarding blood pressure helping with the creation of heart rhythm. Furthermore, it is known that the frequency of the sino-atrial node of the heart is modulated by the autonomic neural and hormonal control and that neural activity is found in respiratory rhythms [24].

One interaction between these two important physiological systems is identified through respiratory sinus arrhythmia (RSA) [25]. Phase synchronization (PS) is also used to describe cardio-respiratory interactions [26]. Studies show that PS and RSA are distinct and not correlated having different physiologic factors affecting them separately. However, both interactions might be used to get a better detection or diagnosis [25].

2.3.1 Respiratory Sinus Arrhythmia

RSA is defined as the HRV in synchronization with respiration and it measures the periodic variation of the heart rate within a breathing cycle [22, 25]. It has this denomination because the heart rhythm is usually called sinus rhythm. This is due to the electrical activity that starts the rhythmic contraction of the heart being originated at the sino-atrial node in the right atrium of the heart [17].

The differences in the RR interval are an indication of the magnitude of RSA. During inspiration, the RR interval is shortened and, during expiration, the RR interval is extended as it is visible in Figure 2.3. This is due to the role of the efferent cardiac vagal nerve whose activity is almost nonexistent during inspiration [22]. RSA increases with cardiac vagal activity and respiratory cycle length. It is caused mainly by central respiratory activity, meaning that inspiration is more valuable [27, 17]. The HR variation is not directly caused by respiration, but it is synchronized with the central rhythm, which projects onto the respiratory muscles [28].

The sympathetic and parasympathetic systems influence the cardiac and respiratory signals. Therefore, both systems also help to generate RSA. Sympathetic activity depends on the activity of neurons that are modulated by respiration. The coupling between this system and respiration is found in multiple nerves. The parasympathetic system causes inhibitory effects that influence the heart rate [21].

Thus, we get two mechanisms of RSA generation: modulation of cardiac vagal neurons by central respiration and inhibition of the efferent cardiac vagal nerve by lung inflation [22].

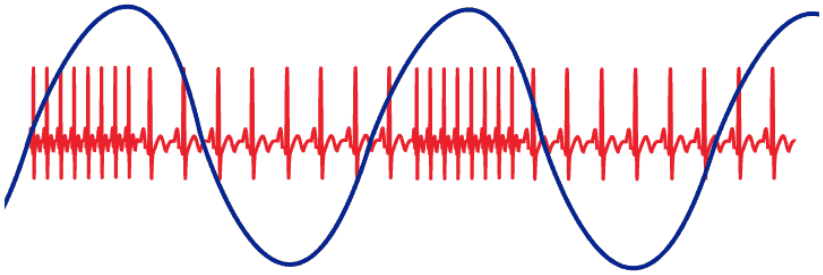


Figure 2.3: Coordination between the HR (ECG represented in red) and the respiratory signal (represented in blue) [29].

RSA can be influenced by factors like age, gender, breathing pattern, or cardiopulmonary function [22]. During daytime, when people are occupied and exercising, RSA is usually not notable, but during rest and sleep it is easy to recognize [30]. RSA is more noticeable in infants than in older people [22]. RSA can also be found in fetus during gestation where parasympathetic activity rules CRC. After birth, the sympathetic system starts to contribute [21]. During the first months of life, CRC needs to develop and stabilize since the body needs to adapt to the conditions outside the uterus [27]. As a person ages, there is a decline in control mechanisms leading to decreased physiologic function and complexity. Therefore, RSA decreases with age [25].

RSA is found during both normal breathing and augmented breathing patterns. Many diseases present cardio-respiratory issues, such as hypertension and heart failure identified by a reduced response in cardiac vagal activity. Thus, RSA being influenced by cardiac vagal activity may help with the detection of certain conditions [21].

2.3.2 Phase synchronization

A common feature of biological oscillatory systems such as the cardiac and respiratory systems is their ability to synchronize despite of their signals having irregular and non-stationary characteristics. Synchronization occurs due to the coupling of non-linear systems [28]. Both noisy and chaotic systems tend to exhibit fluctuations in relative phase and phase slips [24].

Cardio-respiratory PS is known as a consistent occurrence of heartbeats at the same relative phases within consecutive breathing cycles. This is an indication of the temporal organization of the cardiac and respiratory signals. Quantifying PS can help understanding how physiologic regulation affects CRC [25].

It was demonstrated in previous studies [25] that CRC has phase transitions across physiological states, such as the wake/sleep phase and the different sleep stages and will have a specific level of PS during each sleep stage. Besides that, as said in RSA, PS is also known to decrease with age.

Synchronization of heartbeat and respiration in babies occurs in different ratios. It is not well known if PS is essential in the control of CRC. However, it is hypothesized that the lack of synchronization might indicate a pathological condition or that an individual is at risk [24].

2.4 Previous work

A few studies have applied different methods to HRV and CRC to help with the diagnosis of some health conditions.

Reulecke *et al.* at [31] described a study investigating how CRC appears in quiet and active sleep of preterm neonates using time domain, frequency domain and non-linear dynamics analysis, such as mutual information, Poincaré plot analysis and entropy. The results showed that CRC is not entirely developed in very preterm neonates, but in full-term newborns, CRC is clearly present.

Valderas *et al.* at [32] proposed mutual information for the analysis of HRV and interactions between HRV and respiration for emotion recognition as a tool diagnosing psycho-neural illnesses. It was found that mutual information can be used to discriminate different states of emotion.

Hoyer *et al.* at [33] also proposed a study with mutual information to measure HRV and cardio-respiratory interactions after myocardial infarction. The goal in this case was to prove that non-linear methods achieve better results than other methods and the results confirmed the assumption.

Frasch *et al.* at [34] used linear methods and autonomic information flow to quantify cardio-respiratory coordination during quiet and active sleep in full-term babies. The results showed a significant difference in both sleep states, where HR fluctuations are more complex in quiet sleep and CRC is more complex in active sleep.

Zhao *et al.* at [35] analyzed HRV and CRC using multiple entropy-based methods, such as sample entropy and fuzzy measure entropy, to identify different stages of depression. A notable difference in the parameters of CRC between different depression stages was found, as well as significant correlations between entropy measures and depression severity. This led to the conclusion that depression changes the non-linear characteristics of physiological signals.

Pérez *et al.* at [36] proposed a system to identify perinatal hypoxia by analysis of fetal heart rate during labor using NCD. A system with time and frequency domain and moment features was created, achieving the best results with signals from 2 to 4 hours before birth.

Nault *et al.* at [15] used non-linear correlation coefficients and the RSA amplitude to calculate cardio-respiratory coordination. Besides that, time-domain and frequency-domain analysis and sample entropy were used in HRV to assist the detection of LOS. The conclusion was that alterations in HR variability along with systemic inflammation can provide an earlier diagnosis of LOS.

Although many studies have been performed to analyze HRV and CRC, there is a lack of studies that also analyze the RESP signal separately. Additionally, a model that combines some of the mentioned signal processing measures to detect neonatal sepsis hasn't been created and it can be useful to predict this condition and get a treatment earlier without possible sequelae.

Chapter 3

Signal processing methods

For this study, different signal processing methodologies will be applied to physiological signals to help with the faster identification of neonatal sepsis in infants. This chapter is destined to explaining the theory behind the different linear and non-linear methods.

3.1 Time-domain analysis

When data is collected as a function of time, we can say that it is in the time domain. The time-domain analysis is performed on signals without transformation and focuses on identifying the elements' statistical dependence [2].

In the time-domain analysis, there are statistical and geometrical methods. From statistical methods, we can distinguish the mean ($\bar{x}$) and the standard deviation (STD) described in Equations 3.1 and 3.2, respectively, that can be applied to short- or long-term signals. Each point is represented by x_i with $i = 1,...,N$. Another measure that can be calculated is the RMSSD, described in Equation 3.3, representing the square root of the mean of the sum of the squares of the difference between adjacent points [37].

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{3.1}$$

$$STD = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}} \tag{3.2}$$

$$RMSSD = \sqrt{\frac{\sum_{i=1}^{N-1} (x_{i+1} - x_i)^2}{N-1}} \tag{3.3}$$

For HRV [37], the mean reflects sympathetic and parasympathetic activity on the sinus node of the heart. The STD shows the long-term elements responsible for variability. Finally, the RMSSD shows the short cyclical variability of the HR.

Geometrical measurements are based on the histogram of the RR intervals [38]. The most common measure is the HRV triangular index where a triangle with the same area as the histogram is created from the peak of the histogram to the baseline. The index is then calculated as the proportion between the area (total number of intervals) and the height (most frequent interval) of the triangle.

3.2 Non-linear analysis

A biological system is known for its non-linear properties characterized by greater adaptability and low predictability [19]. By analyzing the non-linear behavior of physiological signals, we can better understand their interactions and possibly if they are affected by certain diseases. Studies [34] show that non-linear analysis is more sensitive to physiological autonomic changes and markers of sympathetic and vagal activities.

An important concept in non-linear analysis is complexity, a property that quantifies the amount of structured information. The most common complexity measures are entropy-based and compression-based measures [39].

3.2.1 Entropy-based methods

Shannon introduced the first method for measuring information called entropy. Entropy quantifies the randomness and predictability of a system where higher entropy means a more random and non-ordered system and lower entropy means higher regularity [19]. Equation 3.4 shows the Shannon entropy of a variable X using N values with distribution p_x [39].

$$H(X) = \sum_{i=1}^{N} p_{x_i} \cdot log \frac{1}{p_{x_i}} \tag{3.4}$$

Entropy can be used to quantify the repetition of patterns in physiological signals where a greater regularity can be associated with sickness or aging [19]. Common approaches to calculating entropy are approximate entropy (ApEn) and sample entropy (SampEn).

ApEn was introduced by Pincus and calculates approximately the negative natural logarithm of the conditional probability that, a data set of length N, repeating itself for m points with a tolerance of r, will also repeat itself for m+1 points. Equations 3.5 to 3.7 show how to obtain ApEn [40].

$$C_i^m(r) = \frac{\text{number of j such that } d[x(i),x(j)] \leq r}{N-m+1} \tag{3.5}$$

$$\Phi^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} logC_i^m(r) \tag{3.6}$$

$$ApEn(m,r,N) = \Phi^m(r) - \Phi^{m+1}(r) \tag{3.7}$$

This approach is sensitive to noise, and the data set must be stationary [21]. Besides that, another disadvantage is the dependency on the signal length. For the example of heart rate, a small value of ApEn will tell us that the signal contains repetitive events with minimal fluctuations, which is considered an abnormal behavior [19].

SampEn was introduced by Richman and Moorman to solve some of ApEn's issues and is calculated as the negative natural logarithm of the conditional probability that two sequences are similar for m points will also be similar for m+1 points with a tolerance of r. Equation 3.8 shows how to obtain SampEn [41].

$$SampEn(m,r) = -log\left(\frac{\text{number of pairs having } d[x_{m+1}(i),x_{m+1}(j)] \leq r}{\text{number of pairs having } d[x_m(i),x_m(j)] \leq r}\right) \tag{3.8}$$

This approach is more reliable and less biased than ApEn [18]. SampEn is negatively correlated with regularity and a higher value indicates higher complexity and unpredictability. In other words, a higher value of SampEn indicates that there is less self-similarity in a signal [19]. A disadvantage of this approach is the non-uniform variation of the entropy value with the change of the threshold r [35]. Fuzzy entropy was created to improve the limitation mentioned above, allowing a gradually varied entropy value with the changes of r.

Since we are dealing with complex and irregular signals, a multiscale entropy (MSE) approach was used. This is a measure of the entropy of a signal at different scales, presented in Equation 3.9, being τ the scale factor [42]. For example, scale 1 corresponds to the original time series with length N and scale 2 corresponds to the time series with a length equal to half of N, where each 2 consecutive points are replaced with their average, and so on [42, 43].

For the MSE, five scales were used and the parameters for each signal were $m = 2$ and $r = 0.15$. Then, the scale 1 (MSE_{scale1}) and the sum of all scales (MSE_{sum}) were selected for the final results.

$$y_j^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i \qquad 1 \leq j \leq \frac{N}{\tau} \tag{3.9}$$

Cross-entropy (CE) is used to explore the interaction between two different physiological signals [35]. It can be defined as the possibility of one of the signals (X) being predicted based on the information from the other signal (Y). This approach pays attention to the interaction between the past X and the present Y ignoring the past Y. Another approach to measure cross-entropy is by considering the effects from both series simultaneously focusing on the synchronization between them. Equation 3.10 shows how to calculate the cross-entropy between X and Y.

$$H(X,Y) = \sum_{i=1}^{N} p_{x_i} \cdot log\frac{1}{p_{y_i}} \tag{3.10}$$

Cross-entropy aims to evaluate the interaction between two simultaneous signals under the central autonomic system's influence. This approach aims to measure how synchronous the signals are where a small cross-entropy value is associated with high synchronization.

For this work, CE was calculated in two ways, using the RR signal in the first place and the RESP in the second place ($CE_{RR\text{-}RESP}$), and vice-versa ($CE_{RESP\text{-}RR}$).

3.2.2 Mutual information

Mutual information (MI) [33] is used to describe the amount of information received for a random quantity (η) if another quantity (ξ) is known. Equation 3.11 describes the measure of information H for a discrete random variable (ξ) by a probability distribution (p_m), and Equation 3.12 shows how to obtain the mutual information between 2 variables. Usually, for time-domain signals, the second variable is a shifted version of the first (t and t+τ). The probability distribution corresponds to a division of the signal in a histogram [32].

$$H(\xi) = -\sum_{m} p_m \cdot log(p_m) \tag{3.11}$$

$$MI(\xi,\eta) = H(\eta) - [H(\xi,\eta) - H(\xi)] \tag{3.12}$$

Here, we can distinguish between auto mutual information (AMI) and cross mutual information (CMI) being both based on the Shannon entropy already described. Equation 3.13 and Equation 3.14 represent the AMI function (AMIF) and the CMI function (CMIF), respectively [32]. The first one describes the amount of shared information between the original time series and the time shifted version. The second one describes the amount of shared information between a time series and a different time shifted series. If the signals are independent, AMIF or CMIF is zero, otherwise is positive.

$$AMIF_{xx}(\tau) = H_{x(t)} + H_{x(t+\tau)} - H_{x(t)\cdot x(t+\tau)} \tag{3.13}$$

$$CMIF_{xy}(\tau) = H_{x(t)} + H_{y(t+\tau)} - H_{x(t)\cdot y(t+\tau)} \tag{3.14}$$

Studies show that mutual information can be used to describe the predictability and regularity of signals [32]. Furthermore, AMI can be used to assess the communication of HRV with the ANS and CMI to quantify the interaction between the cardiac and respiratory signals.

To measure the MI between the two signals, the CMI at $\tau = 0$ (MI_0) and the maximum CMI (MI_{max}) were obtained. A 2^6 bin size was selected for the histograms.

3.2.3 Normalized compression distance

The Kolmogorov complexity K(X) of a signal X is the length of the shortest binary program necessary to produce it [36].

Similarity tells us that two objects are similar if the basic components that constitute each one are present in the other one [44]. To measure the similarity between two objects, information distance (ID) is used where two objects are similar if there is a simple way to transform each one of

them into the other one [45]. Equations 3.15 and 3.16 show how to obtain ID where K(X|Y) is the complexity measure to produce X if Y is given and K(Y,X) is the complexity of the concatenation of X and Y [36].

$$K(Y,X) = K(Y) + K(X|Y) = K(X) + K(Y|X) \tag{3.15}$$

$$ID(Y,X) = max\{K(X|Y), K(Y|X)\} \tag{3.16}$$

The problem is that this complexity measure is difficult to use for signals with different sizes. Thus, the idea is to normalize ID in order to obtain an universal measure. The normalized information distance (NID) between two signals is presented in Equation 3.17. However, this measure is only theoretical and cannot be computed [45].

$$NID(X,Y) = \frac{max\{K(X|Y), K(Y|X)\}}{max\{K(X), K(Y)\}} \tag{3.17}$$

To overcome the mentioned problem, NCD is presented as a normalized similarity measure for signals that can be computed [36]. Here, the idea is to get an approximation of K(X) using real-world compressors.

Equation 3.18 shows the computation of NCD for two signals, X and Y, using a real-world compressor Z [45]. The lower its value, the more similar the signals are, the more information is shared, and fewer bits are required to compress them [36].

$$NCD(X,Y) = \frac{Z(XY) - min\{Z(X), Z(Y)\}}{max\{Z(X), Z(Y)\}} \tag{3.18}$$

For two signals, a compressor seeks for information shared between them to reduce the redundancy. If the obtained result is small, there is much information shared between both signals, and, therefore, the signals are similar [44].

Data compression methods depend on *a priori* assumptions about the structure of the data. Usually, there is a two-step approach where the characteristics of the data are discovered and then, with that knowledge, the compression is performed. There is also a one-step approach, known as the adaptative coding, where the encoding scheme changes while the data is being encoded [46].

Several compressors can be used, such as the GZIP compressor that uses the Lempel–Ziv coding [47] and the BZIP2 compressor developed by Julian Seward in 1996 [48]. The Lempel-Ziv algorithm (LZMA) developed in 1977 at [49] is widely used to indicate biomedical signals' randomness [50]. Mahoney at [51] presents multiple compressors like PAQ8, with the best compression rates, and PAQ9A, faster than the previous one.

Here, as in entropy, a multiscale compression (MSC) technique was used. Three different compressors were chosen: GZIP, BZIP2 and LZMA, and five scales were used. Then, the ratio between the compressed and the original signal was calculated for all scales. Finally, for the

results, the scale 1 ratio (MSC_{scale1}) and the sum of all scales' ratios (MSC_{sum}) were selected. For the NCD, the MSC_{scale1} was used with the three compressors.

3.3 Methodologies

For this project, the measures mentioned before were applied to RR and RESP signals with the goal of finding differences between sepsis and control data sets. All the work was developed in programming language using Python and its libraries. Firstly, the time-domain methods, the mean, the STD, and the RMSSD and the non-linear methods, the MSE and the MSC were calculated for both types of signals separately. Then, the joint measures, MI, CE, and NCD, were calculated using both signals to assess the association between signals.

For the linear measures, the mean and STD were already incorporated in the *numpy* Python library. The RMSSD function was created using the equation previously shown. For the non-linear measures, both MSE and MSC functions used were produced in a previous thesis and posteriorly applied to the signals. The MI, CE, and NCD methods did not have a previous computation in Python, so they were designed and tested for this work.

3.4 Statistical analysis

For this work, all measures previously mentioned are continuous and do not present a normal distribution. Therefore, the results were summarized by their median and first and third quartiles, divided by sepsis and control data.

The non-parametric Mann-Whitney rank test was performed to evaluate the significance of the difference between sepsis and control results, where p values of less that 0.05 were considered significant. Significance tests usually aren't enough to fully comprehend the results [52]. When the sample size is large, there is always significance of results. On the other hand, when the sample size is small, even if there are differences between groups, likely, the test would not find them accurately. Therefore, the effect size can be used to measure how meaningful the difference between groups is. Unlike significance tests, effect size doesn't depend on sample size. The larger the effect size, the stronger the importance of the differences between groups. For this work, the effect size by the Cliff's Delta (δ) statistic was used. Table 3.1 shows the interpretation of the effect size values.

Table 3.1: Interpretation of Cliff's Delta values [53].

Cliff's Delta value	**Interpretation**
$\lvert\delta\rvert < 0.147$	Negligible
$0.147 \leq \lvert\delta\rvert < 0.330$	Small
$0.330 \leq \lvert\delta\rvert < 0.474$	Medium
$\lvert\delta\rvert \geq 0.474$	Large

Chapter 4

Time series

For this work, data from sepsis and control data sets was used. This chapter gives a description of the data sets as well as the physiological signals and the pre-processing applied to the signals.

4.1 Data information

We used data from an international study comprising 60 babies of a control data set and 24 babies of a sepsis database. Tables 4.1 and 4.2 show a descriptive analysis of the data for the babies. In the sepsis data set, 10 babies were female, and 14 babies were male. In the control data set, 29 were female, and 31 were male.

Table 4.1: Control data set (N = 60).

	Median (Q1, Q3)
Weight (g)	1125 (950, 1305)
Birth gestational age (days)	201 (193, 208)
Apgar score at 1 min	7 (3.125, 8)
Apgar score at 5 min	9 (8, 10)

Table 4.2: Sepsis data set (N = 24).

	Median (Q1, Q3)
Weight (g)	835 (745, 952.5)
Birth gestational age (days)	192 (182.5, 200.5)
Apgar score at 1 min	8 (5, 10)
Apgar score at 5 min	9 (7, 10)

Figures 4.1 and 4.2 present each baby's gestational age, information about when the signals were collected, and, for the sepsis data set, the exact moment of sepsis detection.

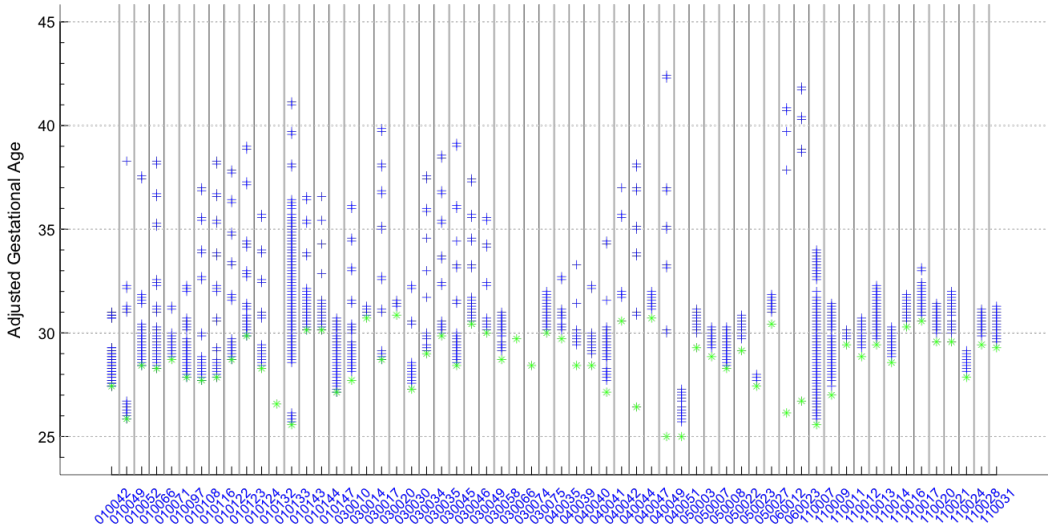


Figure 4.1: Weeks when signals for 60 control babies were collected (the blue + represents each day of signal collecting and the green star the birth gestational age).

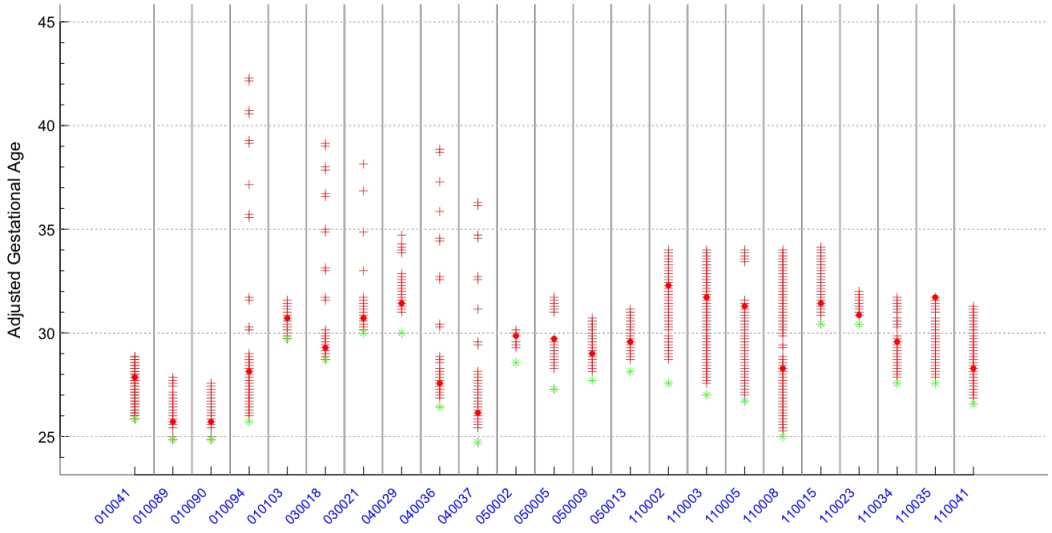


Figure 4.2: Weeks when signals for 23 sepsis babies were collected (the red + represents each day of signal collecting, the green star the birth gestational age and the dot the moment of sepsis detection).

The average $\pm$ STD moment of sepsis detection was 14 $\pm$ 9 days after birth, with an interval for the detection of sepsis from 3 to 33 days after birth.

Babies of the sepsis data set were matched with babies of the control data set by birth gestational age with a maximum difference of $\pm$ 2 days. For every sepsis-control pair, it was checked if the control baby had signals collected at the moment of sepsis of the infected baby. For each neonate, signals were selected at birth, at least 6 hours before sepsis detection, and 3 days after sepsis. At this point, not all babies had a match or signals collected exactly in the needed moments so a smaller sample was used.

For most of the babies, two signals for each type of signal (RR and RESP) were collected. For all these cases and for each measure, the mean between the values of the two signals was calculated.

4.2 Signals' representation and resampling

An example of the signals recorded for an infant is shown in Figure 4.3 with the RR original, the RR filtered and the RESP signals. All the signals were selected during a time frame of thirty minutes. For this work, only the last two were used.

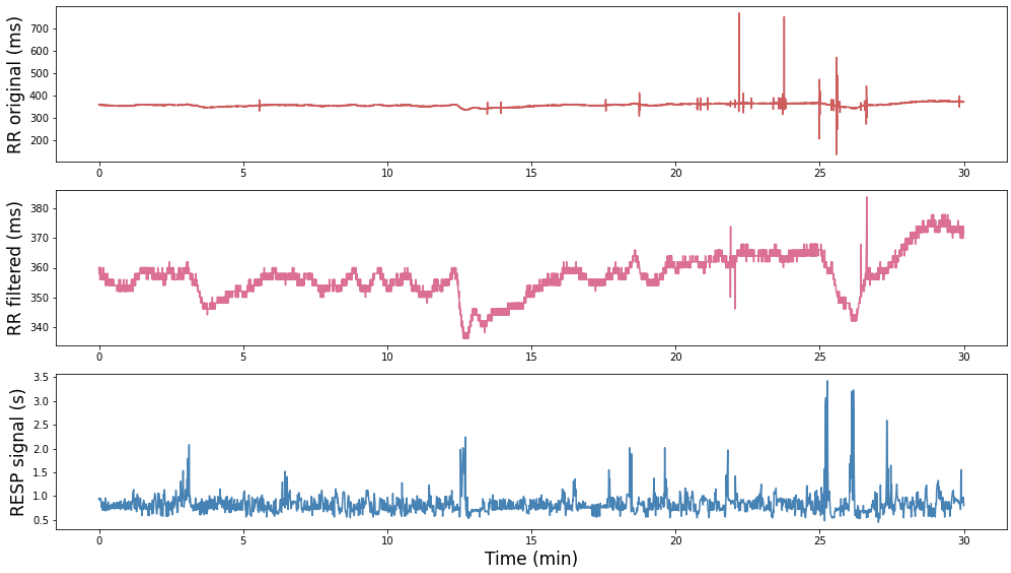


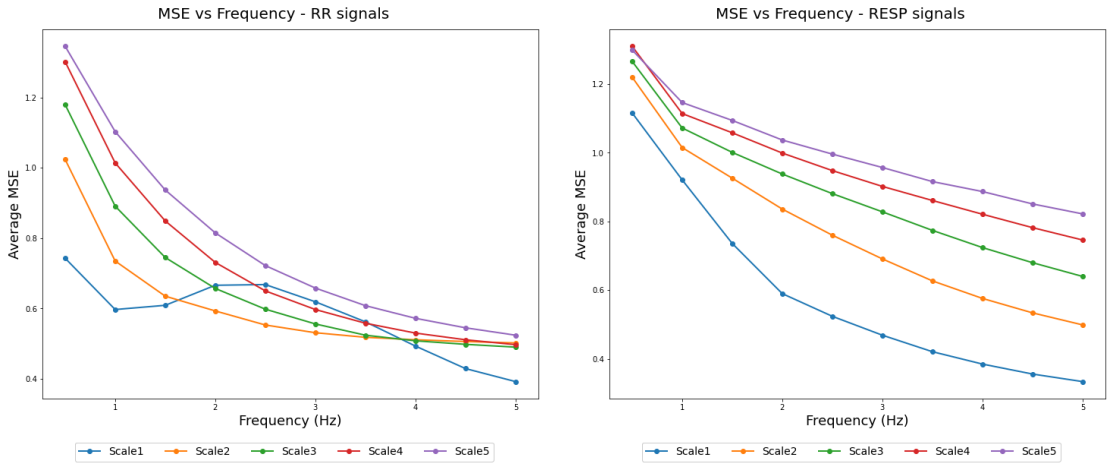
Figure 4.3: Example of the recorded signals.

To apply the non-linear methods to measure the coupling between the signals, they needed to have the same frequency. In this case, the RR signal had a higher frequency (approximately 2.5Hz) than the RESP signal (approximately 1Hz). Therefore, resampling of the signals needed to be performed. For upsampling and downsampling, a function that uses interpolation and the mean between points was applied.

The signals were resampled using several frequencies to evaluate their changes and also how their entropy varies. For this step, we used the signals at birth.

Figures 4.4a and 4.4b show how each scale of entropy changes with frequency for both types of signals. It is visible that as the frequency increases, the entropy decreases and vice-versa.

The chosen frequency was 2Hz since it gave the most similar results in terms of comparison between MSEs. Also, it was a frequency between the original frequencies (2.5Hz and 1Hz). The MSEs for the original and resampled signals are shown in Table 4.3, where small differences were found between most values. The Pearson correlation coefficient (ρ) was used to confirm the correlation between the original and the resampled signals and, as we can see, the large ρ values confirm the similarity between signals.



(a) RR signals. (b) RESP signals.

Figure 4.4: Variation of MSE with different resampling frequencies.

Table 4.3: Average ± STD MSE of original and resampled (2Hz) signals.

Scales	**MSE of RR signals**				**MSE of RESP signals**			
	Original	Resampled	Diff.	ρ	Original	Resampled	Diff.	ρ
1	0.691 ± 0.370	0.666 ± 0.354	0.025	0.949	0.993 ± 0.349	0.590 ± 0.217	0.403	0.816
2	0.558 ± 0.333	0.593 ± 0.313	0.035	0.968	1.034 ± 0.397	0.836 ± 0.315	0.198	0.914
3	0.607 ± 0.318	0.657 ± 0.326	0.050	0.981	1.073 ± 0.404	0.938 ± 0.364	0.135	0.951
4	0.665 ± 0.342	0.731 ± 0.349	0.066	0.987	1.111 ± 0.433	0.999 ± 0.409	0.112	0.961
5	0.736 ± 0.367	0.815 ± 0.384	0.079	0.990	1.125 ± 0.426	1.037 ± 0.416	0.088	0.954

Figures 4.5 and 4.6 show the difference between the original and resampled RR and RESP signals, respectively, for the whole thirty minutes and for a time frame of one minute. This resampling was applied to all the signals for the joint measures' calculation.

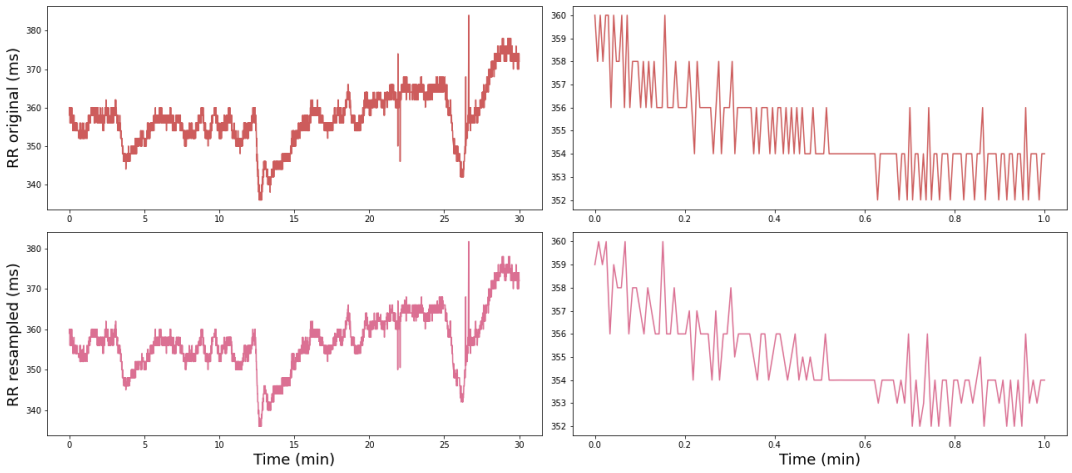


Figure 4.5: Original (2.5Hz) and resampled (2Hz) RR signals.

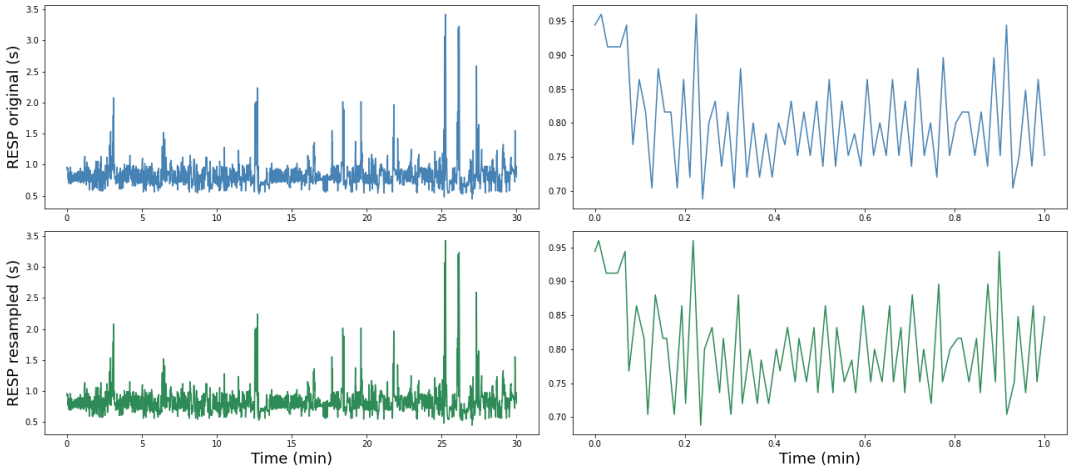


Figure 4.6: Original (1Hz) and resampled (2Hz) RESP signals.

Chapter 5

Results

Several methods were applied to the RR and RESP signals, individually and together, in order to see the differences between sepsis and control babies. The measures were already described in Chapter 3 and this chapter shows the results obtained.

5.1 At birth

The first step was to calculate the mean, STD and RMSSD of all signals at birth, presented in Table 5.1, divided by sepsis and control. Only $RMSSD_{RR}$ showed significantly higher values for control infants when compared with the sepsis ones (large effect size). Furthermore, STD_{RESP} and $RMSSD_{RESP}$ presented a medium effect size.

Table 5.1: Linear measures obtained for the RR and RESP signals at birth. P values lower than 0.05 and medium and large effect sizes are presented in bold.

	Sepsis (n=13) Median (Q1, Q3)	**Control (n=13) Median (Q1, Q3)**	**Mann-Whitney p value**	**Cliff's Delta Effect size**
Mean				
RR	404.93 (394.52, 430.97)	399.55 (386.79, 424.04)	0.644	0.112
RESP	0.98 (0.86, 1.16)	1.04 (0.99, 1.27)	0.356	-0.219
STD				
RR	12.96 (11.94, 18.42)	16.91 (11.02, 21.69)	0.837	-0.053
RESP	0.40 (0.33, 0.73)	0.65 (0.47, 0.82)	0.065	**-0.432**
RMSSD				
RR	3.38 (2.55, 4.86)	5.10 (3.56, 6.00)	**0.036**	**-0.491**
RESP	0.44 (0.38, 0.88)	0.79 (0.49, 1.14)	0.081	**-0.408**

An example of the entropy values for the five scales is displayed in Figure 5.1, for both RR and RESP signals of a baby at birth. The results for MSE_{scale1} and MSE_{sum} are presented in Table 5.2. The results show that there aren't any significant differences between both groups (sepsis and control).

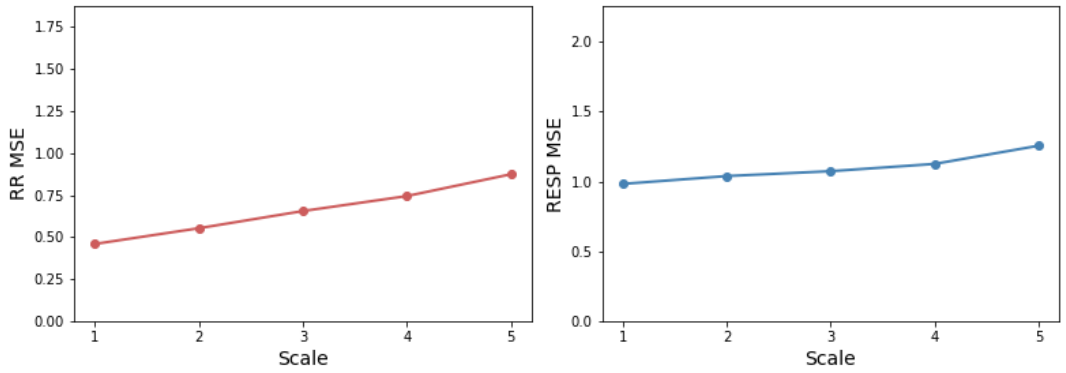


Figure 5.1: Example of the MSE obtained for a baby at birth.

Table 5.2: MSE measures obtained for the RR and RESP signals at birth.

	Sepsis (n=13) Median (Q1, Q3)	**Control (n=13) Median (Q1, Q3)**	**Mann-Whitney p value**	**Cliff's Delta Effect size**
MSE_{scale1}				
RR	0.62 (0.21, 0.71)	0.50 (0.37, 0.62)	0.608	0.124
RESP	0.92 (0.83, 1.37)	1.04 (0.73, 1.27)	0.758	0.077
MSE_{sum}				
RR	3.90 (1.39, 4.12)	3.15 (2.39, 3.66)	0.644	0.112
RESP	4.93 (4.46, 7.0)	5.52 (3.9, 6.43)	0.573	0.136

An example of the ratio values for the five scales of the three MSC compressors is displayed in Figure 5.2, for both types of signals of a baby at birth. From these graphics, it is visible that all three compressors grant similar results.

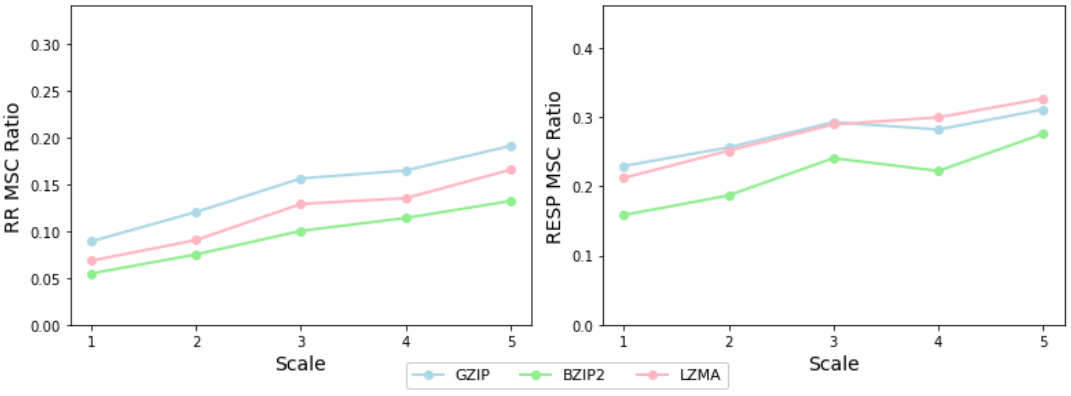


Figure 5.2: Example of the MSC obtained for a baby at birth.

Such as for the MSE, the MSC_{scale1}, the MSC_{sum}, the p value and the effect size are presented in Table 5.3. Since all compressors have a similar behavior (visible in Appendix A), only GZIP is shown in the results. Once again, there aren't any significant differences between both groups at birth.

Table 5.3: MSC measures obtained for the RR and RESP signals at birth.

	Sepsis (n=13) Median (Q1, Q3)	**Control (n=13) Median (Q1, Q3)**	**Mann-Whitney p value**	**Cliff's Delta Effect size**
$GZIP_{scale1}$				
RR	0.08 (0.06, 0.09)	0.08 (0.07, 0.09)	0.412	-0.195
RESP	0.19 (0.19, 0.25)	0.23 (0.21, 0.24)	0.259	-0.266
$GZIP_{sum}$				
RR	0.64 (0.56, 0.72)	0.70 (0.56, 0.73)	0.538	-0.148
RESP	1.23 (1.19, 1.43)	1.36 (1.27, 1.41)	0.238	-0.278

Moving on to the joint non-linear measures, Table 5.4 shows the obtained results. As before, since all compressors exhibit similar results, only GZIP was used here and Appendix A shows the correlation between all NCDs. There aren't any significant differences between both groups at birth.

Table 5.4: Joint measures obtained for the RR and RESP signals at birth.

	Sepsis (n=13) Median (Q1, Q3)	**Control (n=13) Median (Q1, Q3)**	**Mann-Whitney p value**	**Cliff's Delta Effect size**
MI_0	0.30 (0.22, 0.34)	0.28 (0.23, 0.35)	0.505	0.160
MI_{max}	0.34 (0.24, 0.38)	0.31 (0.26, 0.39)	0.538	0.148
$CE_{RR\text{-}RESP}$	6.55 (6.01, 7.53)	6.58 (5.68, 7.55)	0.758	0.077
$CE_{RESP\text{-}RR}$	6.25 (6.14, 6.69)	6.48 (5.38, 7.15)	0.959	-0.018
NCD_{GZIP}	0.25 (0.23, 0.28)	0.27 (0.24, 0.28)	0.473	-0.172

5.2 Near sepsis

Considering the signals obtained near sepsis, Tables 5.5 and 5.6 show the results of the linear and non-linear measures, respectively. Here, the STD_{RR} showed higher significant values for the sepsis group and the $RMSSD_{RR}$ also presented a medium effect size. All non-linear measures presented meaningless differences.

Table 5.5: Linear measures obtained for the RR and RESP signals near sepsis. P values lower than 0.05 and medium and large effect sizes are presented in bold.

	Sepsis (n=19) Median (Q1, Q3)	**Control (n=19) Median (Q1, Q3)**	**Mann-Whitney p value**	**Cliff's Delta Effect size**
Mean				
RR	383.07 (364.47, 404.00)	381.12 (366.19, 397.43)	0.907	0.025
RESP	1.16 (0.97, 1.29)	1.10 (0.99, 1.29)	0.540	0.119
STD				
RR	31.63 (22.17, 43.05)	20.52 (15.34, 30.54)	**0.027**	**0.424**
RESP	0.68 (0.45, 0.78)	0.61 (0.47, 0.74)	0.815	0.047
RMSSD				
RR	4.86 (3.84, 6.18)	3.98 (3.32, 4.66)	0.075	**0.341**
RESP	0.81 (0.53, 0.98)	0.76 (0.55, 0.91)	0.884	0.030

Table 5.6: Non-linear measures obtained for the RR and RESP signals near sepsis.

	Sepsis (n=19) Median (Q1, Q3)	**Control (n=19) Median (Q1, Q3)**	**Mann-Whitney p value**	**Cliff's Delta Effect size**
MSE_{scale1}				
RR	0.37 (0.23, 0.58)	0.54 (0.33, 0.60)	0.465	-0.141
RESP	1.10 (0.89, 1.20)	1.09 (0.90, 1.50)	0.579	-0.108
MSE_{sum}				
RR	2.16 (1.46, 3.08)	3.15 (2.04, 3.63)	0.307	-0.197
RESP	5.74 (4.79, 6.28)	5.66 (4.87, 7.63)	0.770	-0.058
$GZIP_{scale1}$				
RR	0.08 (0.08, 0.09)	0.08 (0.07, 0.09)	0.599	0.102
RESP	0.24 (0.20, 0.24)	0.22 (0.21, 0.24)	0.930	0.019
$GZIP_{sum}$				
RR	0.69 (0.65, 0.72)	0.69 (0.60, 0.73)	0.953	0.014
RESP	1.39 (1.23, 1.42)	1.37 (1.29, 1.43)	0.953	0.014
MI_0	0.24 (0.22, 0.27)	0.23 (0.19, 0.30)	0.579	0.108
MI_{max}	0.27 (0.25, 0.30)	0.25 (0.22, 0.32)	0.335	0.186
$CE_{RR\text{-}RESP}$	5.31 (4.28, 6.56)	5.42 (4.70, 6.42)	0.953	0.014
$CE_{RESP\text{-}RR}$	4.73 (4.22, 6.39)	5.49 (4.55, 6.40)	0.414	-0.158
NCD_{GZIP}	0.26 (0.23, 0.29)	0.26 (0.23, 0.27)	0.748	0.064

Additionally, the differences between the measures at birth and the measures near infection were calculated, to see if there were changes between both moments. Tables 5.7 and 5.8 show the results obtained. Only $RMSSD_{RR}$ showed high significance with a large effect size while $GZIP_{scale1\text{-}RR}$ presented a medium effect size.

Table 5.7: Differences between the linear measures obtained at birth and near sepsis. P values lower than 0.05 and medium and large effect sizes are presented in bold.

	Sepsis (n=12) Median (Q1, Q3)	**Control (n=13) Median (Q1, Q3)**	**Mann-Whitney p value**	**Cliff's Delta Effect size**
Mean				
RR	-15.53 (-65.09, 11.07)	-22.00 (-28.14, 15.90)	0.724	-0.090
RESP	0.12 (-0.02, 0.30)	0.04 (-0.27, 0.25)	0.497	0.167
STD				
RR	10.71 (6.16, 18.62)	8.30 (-2.08, 12.91)	0.289	0.256
RESP	0.02 (-0.09, 0.32)	0.10 (-0.46, 0.26)	0.463	0.179
RMSSD				
RR	1.50 (0.85, 1.76)	-0.30 (-3.93, 0.29)	**0.007**	**0.641**
RESP	0.06 (-0.09, 0.42)	0.05 (-0.54, 0.37)	0.497	0.167

5.3 After sepsis

Finally, this section shows the results obtained for the neonates 3 days after infection, once is considering the time the sepsis babies need to recover after the first take of antibiotics. It is important to mention that, for two babies, the signals were obtained 15 days after infection,

Table 5.8: Differences between the non-linear measures obtained at birth and near sepsis. The medium effect size is presented in bold.

	Sepsis (n=12) Median (Q1, Q3)	**Control (n=13) Median (Q1, Q3)**	**Mann-Whitney p value**	**Cliff's Delta Effect size**
MSE_{scale1}				
RR	-0.02 (-0.25, 0.17)	-0.08 (-0.22, 0.04)	0.605	0.128
RESP	0.04 (-0.26, 0.43)	0.13 (-0.24, 0.45)	0.978	-0.013
MSE_{sum}				
RR	0.22 (-1.87, 0.83)	-0.34 (-1.14, 0.41)	0.644	0.115
RESP	0.19 (-1.23, 1.79)	0.64 (-1.41, 2.31)	0.935	-0.026
$GZIP_{scale1}$				
RR	0.01 (0.00, 0.03)	0.00 (-0.01, 0.01)	0.087	**0.410**
RESP	0.00 (-0.01, 0.04)	0.01 (-0.03, 0.02)	0.644	0.115
$GZIP_{sum}$				
RR	0.03 (-0.02, 0.12)	-0.02 (-0.08, 0.08)	0.314	0.244
RESP	0.01 (-0.06, 0.21)	0.03 (-0.12, 0.08)	0.807	0.064
MI_0	-0.02 (-0.14, 0.03)	-0.07 (-0.08, 0.07)	0.765	-0.077
MI_{max}	-0.02 (-0.14, 0.03)	-0.08 (-0.09, 0.07)	0.849	-0.051
$CE_{RR\text{-}RESP}$	-0.71 (-2.35, 1.32)	-1.75 (-2.66, -0.26)	0.497	0.167
$CE_{RESP\text{-}RR}$	-0.82 (-1.54, 0.19)	-0.41 (-2.55, 0.55)	0.586	0.141
NCD_{GZIP}	0.00 (-0.05, 0.04)	0.00 (-0.03, 0.03)	0.765	0.077

for two babies, 5 days after and, for one baby, 13 days after. The idea here wasn't to obtain a prediction model, but to see differences between the signals near infection and after recovery. It was assumed that all the babies' signals had similar behaviours after recovery, since they were all in their normal state. Therefore, only the difference between measures is presented in Tables 5.9 and 5.10. As visible, no significance was found within the results and all the effect sizes were small or negligible.

Table 5.9: Differences between the linear measures obtained near sepsis and after sepsis.

	Sepsis (n=18) Median (Q1, Q3)	**Control (n=14) Median (Q1, Q3)**	**Mann-Whitney p value**	**Cliff's Delta Effect size**
Mean				
RR	-7.49 (-19.81, 28.02)	11.42 (-12.47, 25.17)	0.556	-0.127
RESP	0.06 (-0.06, 0.14)	0.13 (0.05, 0.42)	0.166	-0.294
STD				
RR	-7.76 (-24.05, -1.16)	-2.18 (-10.10, 1.61)	0.247	-0.246
RESP	0.07 (-0.22, 0.43)	0.15 (-0.01, 0.37)	0.662	-0.095
RMSSD				
RR	0.14 (-2.77, 0.82)	0.05 (-0.97, 1.03)	0.776	-0.063
RESP	0.11 (-0.23, 0.55)	0.21 (-0.02, 0.48)	0.776	-0.063

Table 5.10: Differences between the non-linear measures obtained near sepsis and after sepsis.

	Sepsis (n=18) Median (Q1, Q3)	**Control (n=14) Median (Q1, Q3)**	**Mann-Whitney p value**	**Cliff's Delta Effect size**
MSE_{scale1}				
RR	0.25 (0.06, 0.36)	0.10 (0.03, 0.31)	0.531	0.135
RESP	0.20 (0.01, 0.50)	0.04 (-0.05, 0.43)	0.506	0.143
MSE_{sum}				
RR	1.12 (0.49, 2.04)	1.21 (0.28, 1.51)	0.608	0.111
RESP	0.84 (-0.22, 2.59)	0.06 (-0.23, 2.26)	0.718	0.079
$GZIP_{scale1}$				
RR	0.00 (-0.02, 0.01)	0.00 (0.00, 0.01)	0.531	-0.135
RESP	0.01 (-0.01, 0.02)	0.02 (0.01, 0.04)	0.133	-0.317
$GZIP_{sum}$				
RR	0.02 (-0.07, 0.06)	0.04 (-0.02, 0.07)	0.482	-0.151
RESP	0.04 (-0.05, 0.11)	0.10 (0.02, 0.17)	0.231	-0.254
MI_0	-0.03 (-0.07, 0.07)	-0.02 (-0.04, 0.06)	0.436	-0.167
MI_{max}	-0.05 (-0.08, 0.06)	-0.01 (-0.04, 0.07)	0.262	-0.238
$CE_{RR\text{-}RESP}$	0.23 (-0.29, 0.81)	0.75 (-0.28, 2.59)	0.262	-0.238
$CE_{RESP\text{-}RR}$	1.16 (-0.24, 2.48)	0.07 (-0.8, 0.93)	0.133	0.317
NCD_{GZIP}	0.00 (0.00, 0.02)	0.01 (0.00, 0.03)	0.372	-0.190

Chapter 6

Discussion

In this work, the dynamics of physiological signals and their association was analyzed to help with the earlier detection of neonatal sepsis.

In general, the results showed a lack of differences between the control and sepsis babies, proven by the p values of the Mann-Whitney test and the effect sizes of the Cliff's Delta statistic. These results were expected after recovery and possibly for the signals at birth, but some distinction between both groups was anticipated right before sepsis detection.

Unexpectedly, the time-domain linear measures had a more effective performance than the non-linear measures. For the signals at birth, the $RMSSD_{RR}$ had the biggest significance. Besides that, the STD_{RESP} and the $RMSSD_{RESP}$ showed a medium effect size, which means that the differences between sepsis and control neonates can be notable. For the signals during sepsis, the STD_{RR} had the highest significance and the $RMSSD_{RR}$ showed a medium effect size.

None of the non-linear measures showed good performance in differentiating the two groups of neonates, observable by the large p values and small or negligible effect sizes. This was unexpected since physiological signals tend to have non-stationary behaviour and high complexity.

For the results regarding the association between cardiac and respiratory systems, some coupling was found between the signals.

For MI, if the value is 0, no coupling exists. Otherwise, it is positive and there is information shared between signals [33]. For this work, the MI values obtained were positive but small, indicating some association between RR and RESP signals. For NCD, the lower its values, the more similar the signals are and the fewer bits are necessary to compress them together [36]. Therefore, the small NCD values obtained imply that the RR and RESP signals are similar and, therefore, CRC is present.

On the other hand, CE is negatively correlated with the coupling between signals [35] and, since the values obtained were high, an absence of association was found. A study [31] showed that CRC is not fully developed in the early stages of human life, which could justify these results. Additionally, most of the infants were born very prematurely, with median values of 201 and 192 days, when the normal value is around 280 days (40 weeks) [54]. In an immature system, such

as the case of preterm neonates, the couplings aren't continuous or persistent, being hard to detect [31]. Therefore, it can be assumed that CE isn't the best measure to find those couplings.

Analyzing the differences between measures at birth and near sepsis, most of the measures manifested no relevance for the detection of infection. Since these results were calculated from the previous phases, an absence of significance was already anticipated. Only the $RMSSD_{RR}$ presented a low p value and the $GZIP_{scale1\text{-}RR}$ a medium effect size.

Finally, for the differences between signals before and after sepsis, no significance was found within the results, which is unexpected, once again. After recovery, there should be a difference in the infected neonates' measures while the control babies' measures should remain the same. Signals at different moments of infection could be used in order to find better information.

Throughout this work, several factors could explain the results. Firstly, talking about the data set, the sample size was small, the birth gestational ages were quite different within neonates, and not all babies had signals collected for the three phases. Considering and adjusting for other clinical aspects might be crucial in this database. Secondly, a signals' resampling was performed, which may have altered some patterns in the signals. A different pre-processing by, for instance, adding some filtering could have changed the results. Moreover, naturally, physiological signals present irregular and random patterns, making it difficult to extract information.

Furthermore, only one hour (two 30 minute segments) were considered in each phase for each neonate. Therefore, a more extensive period can allow finding more differences, particularly for the signals near sepsis, where more prominent differences were expected.

Since there were some limitations, more studies should be performed to improve this work and confirm the previous assumptions, with a larger data set, more signals, and different measures.

Chapter 7

Conclusion

Chaos, fractality, and complexity are concepts associated with biological systems. Mathematical formulations that combine these concepts have been applied to physiological signals such as the heart and the respiratory rate to identify and predict possible health conditions.

The goal of this work was to create a model to detect sepsis in neonates using different signal processing methods and accessing the dynamics of the RR and RESP signals. To achieve that, linear methods of time-domain and non-linear methods, such as entropy, compression, MI, CE, and NCD were studied and applied to physiological signals from a sepsis and control database. After that, the values of the measures were compared between the two different groups of infants.

The non-linear measures provided a poor performance in detecting infection in opposition to the linear measures, being RMSSD the measure that showed the best significance and effect size. Furthermore, although the joint measures failed to detect sepsis, MI and NCD performed effectively in CRC analysis. The detection of neonatal sepsis isn't possible with these results, since there isn't enough significance within the values to create a model.

Several other studies could be performed in the future to improve this work, with different and larger data sets and distinct signal processing methods.

References

[1] Anastasia Korolj, Hau-Tieng Wu, and Milica Radisic. A healthy dose of chaos: Using fractal frameworks for engineering higher-fidelity biomedical systems. *Biomaterials*, 219:119363, 2019.

[2] A Eke, P Herman, L Kocsis, and LR Kozak. Fractal characterization of complexity in temporal physiological signals. *Physiological measurement*, 23(1):R1, 2002.

[3] Oliver Faust and Muralidhar G Bairy. Nonlinear analysis of physiological signals: a review. *Journal of Mechanics in Medicine and Biology*, 12(04):1240015, 2012.

[4] Mary Beth Flynn Makic and Elizabeth Bridges. Ce: managing sepsis and septic shock: current guidelines and definitions. *The Leading Voice of Nursing Since 1900| AJN*, 118(2):34–39, 2018.

[5] Lena M Napolitano. Sepsis 2018: definitions and guideline changes. *Surgical infections*, 19(2):117–125, 2018.

[6] Mervyn Singer, Clifford S Deutschman, Christopher Warren Seymour, Manu Shankar-Hari, Djillali Annane, Michael Bauer, Rinaldo Bellomo, Gordon R Bernard, Jean-Daniel Chiche, Craig M Coopersmith, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *Jama*, 315(8):801–810, 2016.

[7] Andi L Shane, Pablo J Sánchez, and Barbara J Stoll. Neonatal sepsis. *The lancet*, 390(10104):1770–1780, 2017.

[8] Birju A Shah and James F Padbury. Neonatal sepsis: an old problem with new insights. *Virulence*, 5(1):170–178, 2014.

[9] Joanna Barnden, Vanessa Diamond, Paul Anthony Heaton, and Siba Prosad Paul. Recognition and management of sepsis in early infancy. *Nursing children and young people*, 28(10), 2016.

[10] Muhammed Ershad, Ahmed Mostafa, Maricel Dela Cruz, and David Vearrier. Neonatal sepsis. *Current emergency and hospital medicine reports*, 7(3):83–90, 2019.

[11] Susan L Lukacs and Stephanie J Schrag. Clinical sepsis in neonates and young infants, united states, 1988-2006. *The Journal of pediatrics*, 160(6):960–965, 2012.

[12] Eman M Rabie Shehab El-Din, Mohamed M Adel El-Sokkary, Mohamed Reda Bassiouny, and Ramadan Hassan. Epidemiology of neonatal sepsis and implicated pathogens: a study from egypt. *BioMed research international*, 2015, 2015.

[13] Faith Kim, Richard A Polin, and Thomas A Hooven. Neonatal sepsis. *British Medical Journal*, 371, 2020.

[14] Xia Qiu, Li Zhang, Yu Tong, Yi Qu, Huiqing Wang, and Dezhi Mu. Interleukin-6 for early diagnosis of neonatal sepsis with premature rupture of the membranes: A meta-analysis. *Medicine*, 97(47), 2018.

[15] Stéphanie Nault, Vincent Creuze, Sally Al-Omar, Annabelle Levasseur, Charlène Nadeau, Nathalie Samson, Roqaya Imane, Sophie Tremblay, Guy Carrault, Patrick Pladys, et al. Cardiorespiratory alterations in a newborn ovine model of systemic inflammation induced by lipopolysaccharide injection. *Frontiers in physiology*, 11:585, 2020.

[16] Gernot Ernst. Heart-rate variability—more than heart beats? *Frontiers in public health*, 5:240, 2017.

[17] Michael Schiek, Friedhelm R Drepper, Ralf Engbert, H-H Abel, and K Suder. Cardiorespiratory synchronization. In *Nonlinear analysis of physiological data*, pages 191–209. Springer, 1998.

[18] Trang Nguyen Phuc Thu, Alfredo I Hernández, Nathalie Costet, Hugues Patural, Vincent Pichot, Guy Carrault, and Alain Beuchée. Improving methodology in heart rate variability analysis for the premature infants: Impact of the time length. *PloS one*, 14(8):e0220692, 2019.

[19] Dimitriy A Dimitriev, Elena V Saperova, and Aleksey D Dimitriev. State anxiety and nonlinear dynamics of heart rate variability in students. *PloS one*, 11(1):e0146131, 2016.

[20] Hongzu Li and Pierre Boulanger. A model-based approach for arrhythmia detection and classification. *Lecture Notes in Computer Science*, page 429–436, 2018. URL: `https://link.springer.com/chapter/10.1007/978-3-030-04375-9_37`, `doi:10.1007/978-3-030-04375-9_37`.

[21] Alfredo J Garcia III, Jenna E Koschnitzky, Tatiana Dashevskiy, and Jan-Marino Ramirez. Cardiorespiratory coupling in health and disease. *Autonomic Neuroscience*, 175(1-2):26–37, 2013.

[22] Fumihiko Yasuma and Jun-ichiro Hayano. Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm? *Chest*, 125(2):683–690, 2004.

[23] Rosemary SC Horne. Cardio-respiratory control during sleep in infancy. *Paediatric respiratory reviews*, 15(2):163–169, 2014.

[24] Ralf Mrowka, Andreas Patzak, and Michael Rosenblum. Quantitative analysis of cardiorespiratory synchronization in infants. *International Journal of Bifurcation and Chaos*, 10(11):2479–2488, 2000.

[25] Ronny P Bartsch, Aicko Y Schumann, Jan W Kantelhardt, Thomas Penzel, and Plamen Ch Ivanov. Phase transitions in physiologic coupling. *Proceedings of the National Academy of Sciences*, 109(26):10181–10186, 2012.

[26] Ronny P Bartsch, Kang KL Liu, Qianli DY Ma, and Plamen Ch Ivanov. Three independent forms of cardio-respiratory coupling: transitions across sleep stages. In *Computing in Cardiology 2014*, pages 781–784. IEEE, 2014.

[27] Maristella Lucchini, William P Fifer, Manuela Ferrario, and Maria G Signorini. Feasibility study for the assessment of cardio-respiratory coupling in newborn infants. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5509–5512. IEEE, 2016.

[28] Carsten Schäfer, Michael G Rosenblum, Hans-Henning Abel, and Jürgen Kurths. Synchronization in the human cardiorespiratory system. *Physical Review E*, 60(1):857, 1999.

[29] Eric Morgan. All about hrv part 4: Respiratory sinus arrhythmia – mindware technologies support, Sep 2017. URL: `https://support.mindwaretech.com/2017/09/all-about-hrv-part-4-respiratory-sinus-arrhythmia/`.

[30] Thomas Penzel, Jan W Kantelhardt, Ronny P Bartsch, Maik Riedl, Jan F Kraemer, Niels Wessel, Carmen Garcia, Martin Glos, Ingo Fietze, and Christoph Schöbel. Modulations of heart rate, ecg, and cardio-respiratory coupling observed in polysomnography. *Frontiers in physiology*, 7:460, 2016.

[31] Sina Reulecke, Steffen Schulz, and Andreas Voss. Autonomic regulation during quiet and active sleep states in very preterm neonates. *Frontiers in physiology*, 3:61, 2012.

[32] María Teresa Valderas, Juan Bolea, Pablo Laguna, Raquel Bailón, Montserrat Vallverdú, et al. Mutual information between heart rate variability and respiration for emotion characterization. *Physiological measurement*, 40(8):084001, 2019.

[33] Dirk Hoyer, Uwe Leder, Heike Hoyer, Bernd Pompe, Michael Sommer, and Ulrich Zwiener. Mutual information and phase dependencies: measures of reduced nonlinear cardiorespiratory interactions after myocardial infarction. *Medical engineering & physics*, 24(1):33–43, 2002.

[34] MG Frasch, U Zwiener, D Hoyer, and M Eiselt. Autonomic organization of respirocardial function in healthy human neonates in quiet and active sleep. *Early human development*, 83(4):269–277, 2007.

[35] Lulu Zhao, Licai Yang, Zhonghua Su, and Chengyu Liu. Cardiorespiratory coupling analysis based on entropy and cross-entropy in distinguishing different depression stages. *Frontiers in physiology*, 10:359, 2019.

[36] Óscar Barquero-Pérez, Ricardo Santiago-Mozos, José M Lillo-Castellano, Beatriz García-Viruete, Rebeca Goya-Esteban, Antonio J Caamaño, José L Rojo-Álvarez, and Carlos Martín-Caballero. Fetal heart rate analysis for automatic detection of perinatal hypoxia using normalized compression distance and machine learning. *Frontiers in physiology*, 8:113, 2017.

[37] H Nagendra, Vinod Kumar, and Shaktidev Mukherjee. Cognitive behavior evaluation based on physiological parameters among young healthy subjects with yoga as intervention. *Computational and mathematical methods in medicine*, 2015, 2015.

[38] A Voss, V Baier, S Schulz, and KJ Bar. Linear and nonlinear methods for analyses of cardiovascular variability in bipolar disorders. *Bipolar disorders*, 8(5p1):441–452, 2006.

[39] Teresa Henriques, Hernâni Gonçalves, Luís Antunes, Mara Matias, João Bernardes, and Cristina Costa-Santos. Entropy and compression: two measures of complexity. *Journal of Evaluation in Clinical Practice*, 19(6):1101–1106, 2013.

[40] Steven M Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6):2297–2301, 1991.

[41] Joshua S Richman, Douglas E Lake, and J Randall Moorman. Sample entropy. In *Methods in enzymology*, volume 384, pages 172–184. Elsevier, 2004.

[42] Madalena Costa, Ary L Goldberger, and C-K Peng. Multiscale entropy analysis of biological signals. *Physical review E*, 71(2):021906, 2005.

[43] Teresa Henriques, Maria Ribeiro, Andreia Teixeira, Luísa Castro, Luís Antunes, and Cristina Costa-Santos. Nonlinear methods most applied to heart-rate time series: A review. *Entropy*, 22(3):309, 2020.

[44] Manuel Cebrián, Manuel Alfonseca, Alfonso Ortega, et al. Common pitfalls using the normalized compression distance: What to watch out for in a compressor. *Communications in Information & Systems*, 5(4):367–384, 2005.

[45] Paul MB Vitányi, Frank J Balbach, Rudi L Cilibrasi, and Ming Li. Normalized information distance. In *Information theory and statistical learning*, pages 45–82. Springer, 2009.

[46] Gordon V. Cormack and R Nigel S Horspool. Data compression using dynamic markov modelling. *The Computer Journal*, 30(6):541–550, 1987.

[47] Jean-Loup Gailly and Mark Adler. The gzip home page, 2021. URL: `http://www.gzip.org`.

[48] Julian Seward. Bzip2, 2021. URL: `http://www.bzip.org/`.

[49] Jacob Ziv and Abraham Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on information theory*, 23(3):337–343, 1977.

[50] Jing Hu, Jianbo Gao, and Jose C Principe. Analysis of biomedical signals by the lempel-ziv complexity: the effect of finite data size. *IEEE Transactions on Biomedical Engineering*, 53(12):2606–2609, 2006.

[51] Matt Mahoney. Data compression programs, 2013. URL: `http://mattmahoney.net/dc/`.

[52] Gail M Sullivan and Richard Feinn. Using effect size—or why the p value is not enough. *Journal of graduate medical education*, 4(3):279, 2012.

[53] Zhiyuan Wan, Xin Xia, David Lo, and Gail C Murphy. How does machine learning change software development practices? *IEEE Transactions on Software Engineering*, 2019.

[54] Anne Marie Jukic, Donna D Baird, Clarice R Weinberg, D Robert McConnaughey, and Allen J Wilcox. Length of human pregnancy and contributors to its natural variation. *Human reproduction*, 28(10):2848–2855, 2013.

Appendix A

Results: Correlation between compressors

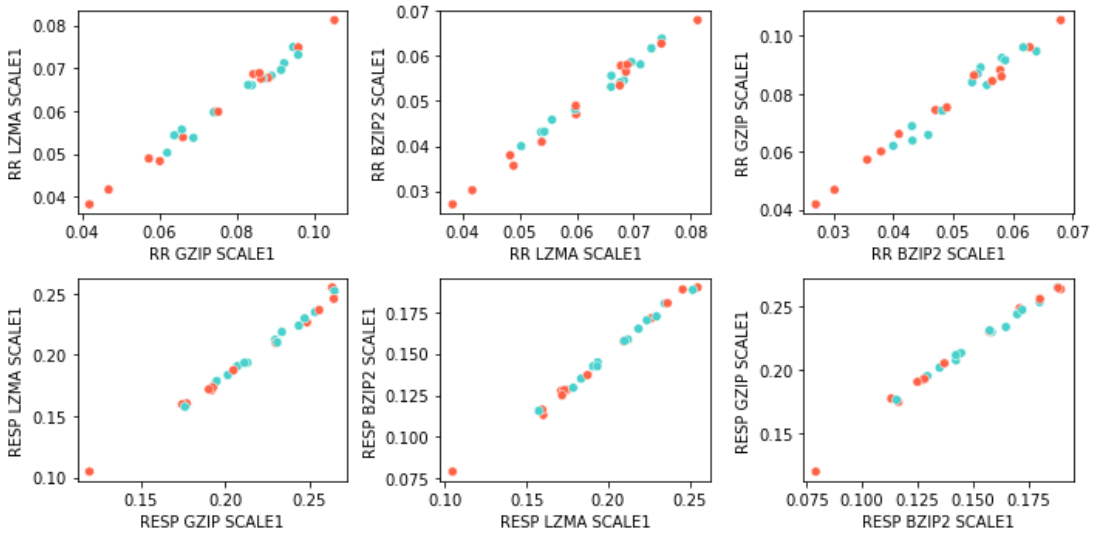


Figure A.1: Correlation between the MSC_{scale1} of the three compressors at birth. All of them present a linear behavior.

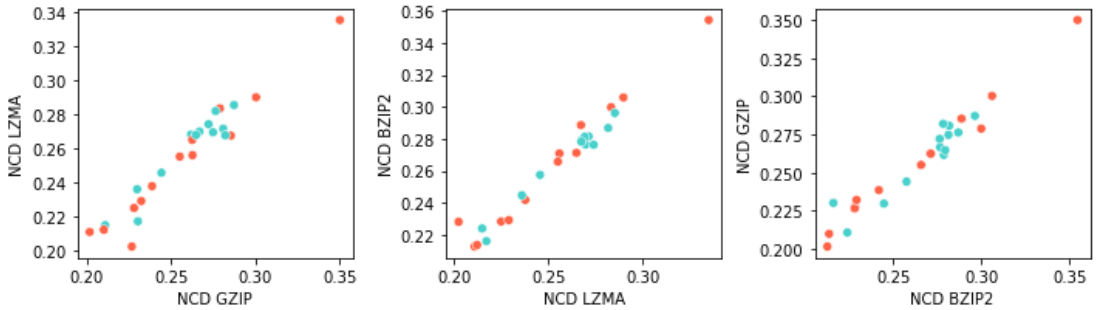


Figure A.2: Correlation between the NCD of the three compressors at birth. All of them present a linear behavior.