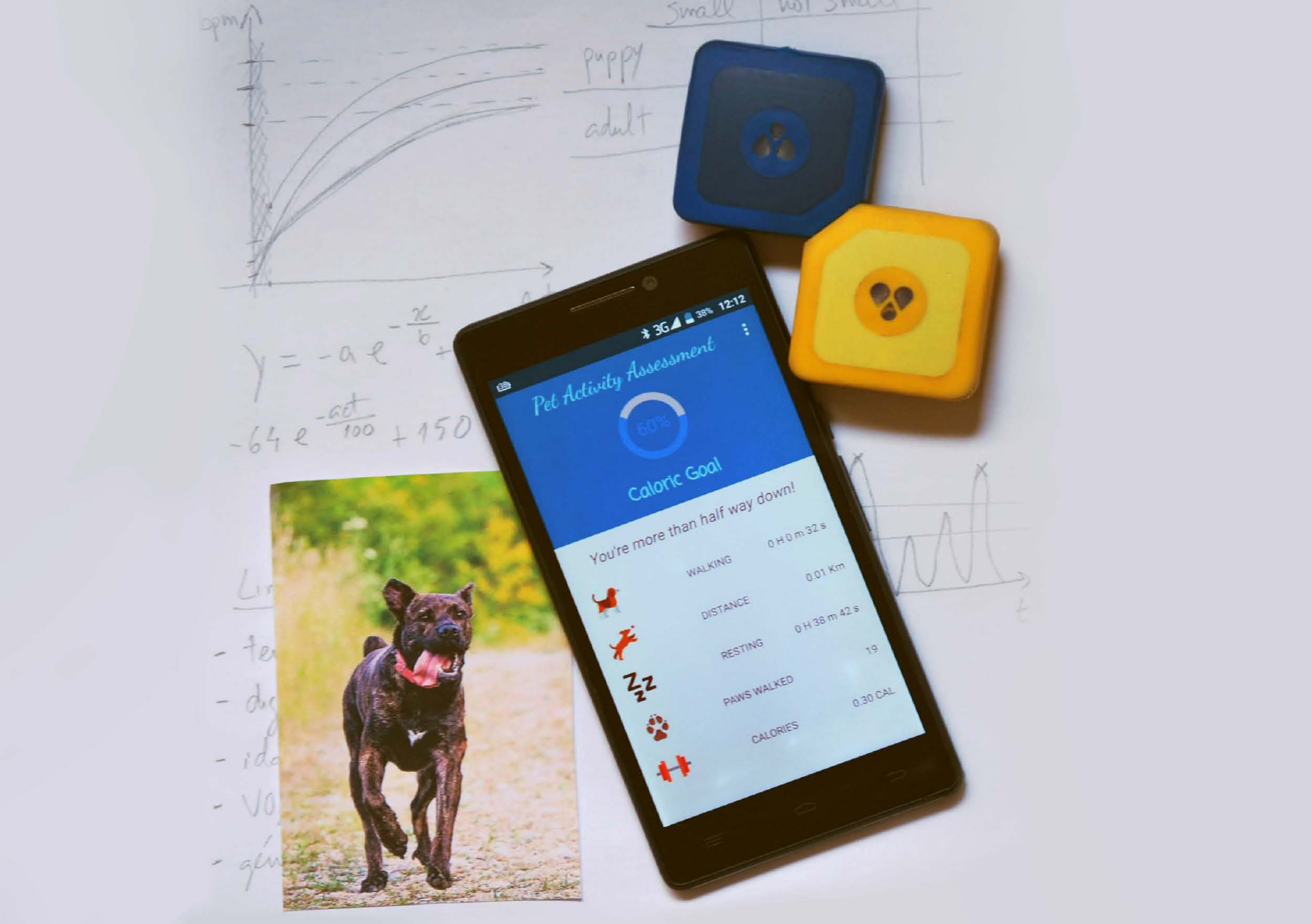


João Luís Barreto Garcia de Sousa

PET ANIMAL ACTIVITY ASSESSMENT USING THREE-DIMENSIONAL ACCELEROMETRY

A thesis submitted in fullfillment of the requirements
for the degree of Master of Science of Biomedical Enginnering.

July 2017

· U C ·

Universidade de Coimbra

Pet animal activity assessment using three-dimensional accelerometry

by
João Luís Barreto Garcia de Sousa

Under the supervision of

João Cardoso, Universidade de Coimbra
André Carvalheira, Findster

A thesis submitted in fulfilment of the requirements for the degree of Master of Science of Biomedical Engineering

Departamento de Física
Universidade de Coimbra
7 Julho 2017

This work was developed in collaboration with:

FINDSTER

Findster Technologies
Porto, Portugal

LIBPhys-UC

Laboratory for Instrumentation, Biomedical Engineering and Radiation Physics (LIBPhys)
University of Coimbra
Coimbra, Portugal

Para os meus pais, avós e tio, que sempre me fizeram saltar mas nunca deixaram cair...

Acknowledgments

Findados estes cinco anos de percurso académico, torna-se importante agradecer a todas as pessoas que dele fizeram parte e, direta ou indiretamente, me ajudaram a completar uma grande e especial etapa da minha vida. A essas pessoas e particularmente aos meus amigos, um muito obrigado.

Obrigado à Findster e especialmente aos que dela fazem parte, pelo agradável espírito que proporcionam a quem integra a equipa e pela paciente transmissão de conhecimento.

Desta equipa destacar o meu obrigado ao Eng. Carlos Pereira por ser parte do corpo orientador deste projeto, pelo seu pragmatismo e experiência que me permitiram ultrapassar obstáculos doutra forma (ainda mais) complexos.

Obrigado ao Eng. André Carvalheira pela constante disponibilidade e mais ainda pela sua genuína preocupação. A sua orientação foi em larga escala um dos fatores que permitiram a este projeto ser o que é hoje.

Obrigado ao Professor João Cardoso que sempre acompanhou de perto o decorrer deste projeto, sendo capaz de dar importantes conselhos em partes críticas do trabalho aliadas a um olhar atento para pormenores de outra forma subvalorizados.

Obrigado ao Nuno Magalhães, Helena Lei, Cristina Ribeiro, João Pedro juntamente com a sua equipa da "Mania dos cães" e a todas as pessoas que gentilmente e sem qualquer interesse cederam os seus amados cães para a realização de testes que nem sempre se apresentaram fáceis.

Um especial obrigado ao Dr. Miguel Faria do hospital veterinário da Universidade do Porto que demonstrou interesse e crença neste projeto desde o seu início, tendo sido o seu contributo crucial para o desenrolar do mesmo.

Com o mais importante obrigado, agradeço à minha família que, sempre presente nos bons e maus momentos foram não só uma referência mas um suporte que nunca me deixou nem deixará cair. Tudo o que sou e faço devo-o a vocês.

Finalmente um carinhoso agradecimento e reconhecimento à "Becas" pela sua extrema cooperação e paciência em todos os testes e repetições, que permitiram o ajuste e desenvolvimento do sistema apresentado por este projeto.

Special recognition to:

UP VET

U.PORTO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

"The most certain way to succeed is always to try just one more time."

Thomas A. Edison

Abstract

Worldwide population is becoming increasingly sedentary and obese. Similar concerns are being extended to companion animals whose incidence in countries like the United States has surpassed half of the entire pet dog's population. As the most common pets' disorder, obesity is known to predispose a range of serious medical conditions, meaning that its importance should not be underestimated. As in humans, dog obesity has two cornerstones, an inadequate dietary regime and the lack of physical activity, existing on the latter a lack of quantitative and pragmatic tools to assess energy expenditure during animal exercise and the consequent impact on weight loss.

This work aims to build an algorithm capable of both qualifying and quantifying dog physical activity while only using a 3-axis accelerometer. The qualification is made into one of five pre-defined levels (stop, walk, trot, run, intense run) and the quantification by converting the activity into a number of calories.

That is accomplished by the creation of two main features, being the first a canine pedometer capable of accounting the number of steps given by the animal in any environment and the second a relation between acceleration and spent calories.

Keywords: Weight Management, Accelerometer, Dog Activity Tracker, Heart Rate, Energy Expenditure

Resumo

A população mundial está a tornar-se cada vez mais sedentária e obesa. Preocupações semelhantes estão a ser alastradas a animais de estimação cuja incidência em países como os Estados Unidos já ultrapassou metade da população total de animais caninos. Sendo a doença mais comum em animais de estimação, obesidade é conhecida pela sua predisposição a uma série de condições médicas graves, significando que a sua importância não deve ser subestimada. Tal como em humanos, obesidade canina assenta em dois fatores principais, um regime alimentar inadequado e a falta de exercício físico, sendo que nesta última existe uma falta de ferramentas pragmáticas e quantitativas para o gasto de energia durante exercício físico e o consequente impacto na perda de peso.
Este projeto aponta para a criação de um algoritmo capaz de qualificar e quantificar atividade física usando apenas um acelerómetro de 3 eixos. A qualificação é feita num de cinco níveis pré-definidos (stop, walk, trot, run, intense run) e a quantificação pela conversão dessa atividade em número de calorias.
O mesmo é alcançado com a criação de duas características principais, sendo a primeira um pedómetro canino capaz de contar o número de passos dado pelo animal em qualquer ambiente e a segunda uma uma relação entre aceleração e calorias gastas.

Palavras-Chave: Gestão de peso, Acelerómetro, Rastreador de Atividade Canina, Frequência Cardíaca, Gasto de Energia

Abbreviations

AAHA American Animal Hospital Association

APOP Association for Pet Obesity Prevention

BCS Body Condition Score

BEE Basal Energy Expenditure

BLE Bluetooth Low Energy

BMI Body Mass Index

BMR Basal Metabolic Rate

BW Body Weight

EAT Exercise Associated Thermogenesis

EE Energy Expenditure

FFT Fast Fourier Transform

FIR Finite Impulse Response

GATT General Attribute Profile

GPS Global Positiong System

IIR Infinite Impulse Response

IoT Internet of Things

KNN K Nearest Neighbour

LED Light Emitting Diode

MA Mean Acceleration

MAD Mean Absolute Deviation

MEMS Micro-ElectroMechanical Systems

MER Maintenance Energy Requirements

MET Metabolic Equivalent Task

MI Movement Intensity

NEAT Non-Exercise Associated Thermogenesis

RER Resting Energy Requirements

RF RadioFrequency

RQ Respiratory Quotient

RSA Respiratory Sinus Arrhytmia

SMA Signal Magnitude Area

SNR Signal to Noise Ratio

SSE Sum of Squared Errors

SVM Signal Magnitude Vector

TDEE Total Daily Energy Expenditure

UC University of Coimbra

UUID Universally Unique Identifier

WHO World Health Organization

List of Figures

List of Tables

Contents

Chapter 1

Introduction

1.1 Context

Obesity, defined as an abnormal accumulation of body fat above the ideal weight is an escalating global problem both in humans and companion animals and one of the most common health problems of today's society [1]. Overweight and obesity are two terms often used to describe the disorder but they show two different levels of the abnormality, where the simple weight-for-height index BMI or Body Mass Index is used to separate these two levels defined as:

$$\mathrm{BMI} = \frac{\mathrm{weight(Kg)}}{\mathrm{height}^2(\mathrm{m}^2)} \tag{1.1}$$

A person considered to be overweight has a BMI equal or greater than 25 (kg/m^2) while an obese individual has a BMI greater than 30. Considering worldwide data, obesity has more than doubled since 1980 and in 2014 more than 1.9 billion adults were overweight where 600 million of these, obese. Moreover, 41 million children under the age of 5 were overweight at the same year [2]. Data from the World Health Organization (WHO) claim that, for example, in Portugal 59% of the adult population is overweight and 24% is obese. The forecasts are a proof that if no measures are taken, the male percentage of obesity will increase from 22% to 27%. Cases like the United States of America show even worse estimates where more than 2 in 3 adults are considered to be overweight or obese [3] where one of the main reasons for this to happen is the common citizen sedentary routine supported by the lack of physical activity and a bad dietary regime. In the United States alone 60% of its population owns a pet [4] and these kind of habits are transmitted, giving rise to obesity problems in companion animals.

Even considering that there is not as much information available for this matter as there is for humans, a considerable amount of research has been made, and it was possible to find out that the percentage of overweight pets range from 22% to 40%, making obesity the most common dog nutritional disorder [5].
As for humans pet obesity has two cornerstones:

- Inadequate dietary regime, usually represented by an excessive dietary intake.
- Lack of physical activity, usually disseminated from the pet owner resulting in an inadequate use of energy, with a very high risk factor for a large number of other diseases that reduce the longevity and quality of life of these animals.

However the key factor to be considered in this issue is that obesity stands as a preventable problem [2] and since the pet/owner connection is becoming increasingly significant throughout generations, it might be crucial to address the problem of obesity in companion animals since it is an illness that can affect both the animal and its owner.

1.2 Technology's Role

Technology can be defined as the branch of knowledge that deals with the creation and use of technical means and their interrelation with life, society, and the environment, drawing upon such subjects as industrial arts, engineering, applied science, and pure science [6].

In fact it is indisputable that technology has an increasing role in our daily lives, with an influence on diverse domains such as industry, with the automation of production lines, consequence of a big industrial revolution on the 19th century; communication, with the constant presence of the world wide web where drafting and sending a message as well as receiving an instant reply can be done easily, putting aside the distant reality of manual mail writing; education, where the traditional presential teacher/student system has been replaced in some cases by online platforms of distance e-learning; health, with the development of important areas such as biomedical engineering, responsible for establishing a bridge between the technical and the medical sectors incorporated by equipments such as X-ray imaging equipment, magnetic ressonance, and many others; fitness, with a wide set of gadgets both eletronic such as smart watches, smart scales, headphones, or embedded into clothing pieces like wrist/arm bands, all of them composed by a set of sensors capable of measuring different signals responsible for a more efficient training. Also the market of mobile applications for mobile devices is a significant part of technology in this domain, existing a tremendous diversity in this field, with applications for every taste. Complementing the stated domains, technology is also present in others like agriculture, transportation, banking, purchasing, among others.

If the domain of physical activity/fitness is to be focused, it is relevant to higlight that most of the technologies applied to this field are based on the optimized use of sensors, where a sensor is a device that detects and responds to some type of input such as light, heat, motion, pressure, or other from a great number of other environmental phenomena.

1.2.1 Wearable Sensors

This subset of the whole wide group of sensors, often refered to as wearable technology, wearable devices or even wearables are meant to be electronic technologies incorporated into clothing parts or accessories which can be comfortably worn.

As it can be seen in figure 1.1 examples of wearable devices include watches, glasses, e-textiles, headbands, bracelets, hearing devices and even some invasive options like implanted micro-chips. In all these examples the basic principle is the same, monitor the health and well being of the users by obtaining vital signs and reporting biofeedback information in real time [8]. Although the potential of this technology exceeds the health and fitness domain with the application in gaming and entertainment when combined with augmented reality, when focusing on the fitness field the principles applied are constant. A wearable sensor is expected to measure a signal that can be related with motion (3-axis, pressure sensors placed in shoes, 9-axis inertial measurement unit), heat (with a thermometer) or heart rate (with use of photopletismography).

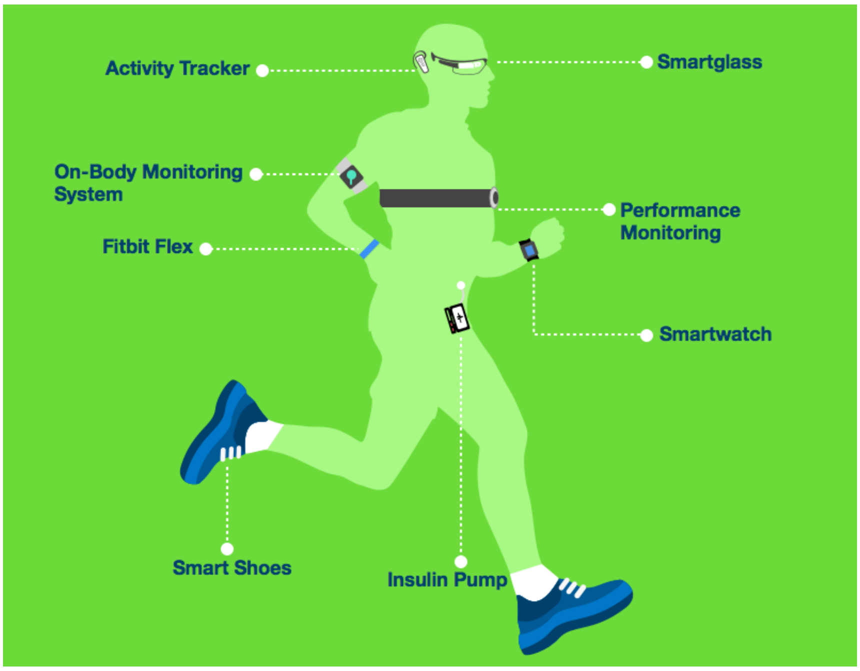


Figure 1.1: Set of wearable devices [7]

The main goal of the described group of sensors is to obtain these signals in the best way possible so that other components of the device can process them and apply robust algorithms capable of giving relevant information such as lost calories during an exercise, heart rate, heart rate variability, etc.

1.2.2 Internet of Things (IoT)

Alongside wearable sensors, Internet of Things triggers even better systems since nowadays everything is connected and sharing information constantly.

Internet of things can be defined as a system of interrelated computing devices provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction [9]. Even though the concept has been developed for decades, evolving from the confluence of wireless technologies, microelectromechanical systems (MEMS) and the internet, it remained unnamed until 1999.

Activity trackers like *FitBit* or *Garmin*, fitness applications and smart sportswear are good examples of the described concept where different sensors monitor physical activities and communicate (often via *Bluetooth*) to a smartphone or to the web for purposes such as in-cloud data storage for easy sharing. This way the user is empowered with an end-to-end solution, with data visualization across all devices. Since monitoring is one of the most important components in a weight loss strategy [5], using concepts like this one towards obesity, is a good way of not only tackling the problem, but also preventing it from happening at all.

1.3 Motivation

As refered earlier, tackling company animals' obesity problem can increase their longevity and life quality translating into a more satisfied and fullfilled owner (and animal).

1.3.1 Obesity Diagnosis

The diagnosis process of an obese (or overweight) animal, based on the principle of body composition assessment, is not an easy task as it is in humans. Many methods can be used such as:

- **X-ray absorptiometry** - consisting of exposing the animal to X-ray beams with different energy levels and assess the absorption in the tissues;
- **Ultrasound** - where the animal is subjected to a beam of high frequency sound waves that reflect upon the body structures and, after their return, make a piezo-electric crystal vibrate, where the time between the sent and received sound wave interpreted to generate an image;
- **Magnetic Ressonance Imaging** - technique used in humans that takes advantage of the nuclear properties of the hydrogen atoms in the patient to render an image that distinguishes the adipocytes from the remaining tissue;

These techniques are tremendously complex, not only from the exam point of view but also logistically, if we consider that the animal must be immobilized, most of the times with use of anesthetics, professional staff must be involved to perform the exam, etc. Moreover the cost of such an alternative is for most pet owners unbearable forcing them to give up on a way to diagnose or simply keep track of their pets' health. Another alternative to these complex exams are quantitative procedures like morphometry, that refer to a variety of measured parameters such as skinfold thickness, head and thorax length, used to estimate body composition that while simpler than the previous methods, still represent a complex evaluation, requiring a physician or a veterinarian each time a measurement is made.

To simplify this issue, some owners and the majority of veterinarians make use of subjective methods like body condition scoring or BCS, a multi-level scheme that assess visual and palpable characteristics that correlate to subcutaneous, abdominal and superficial musculature as the scheme in figure 1.2

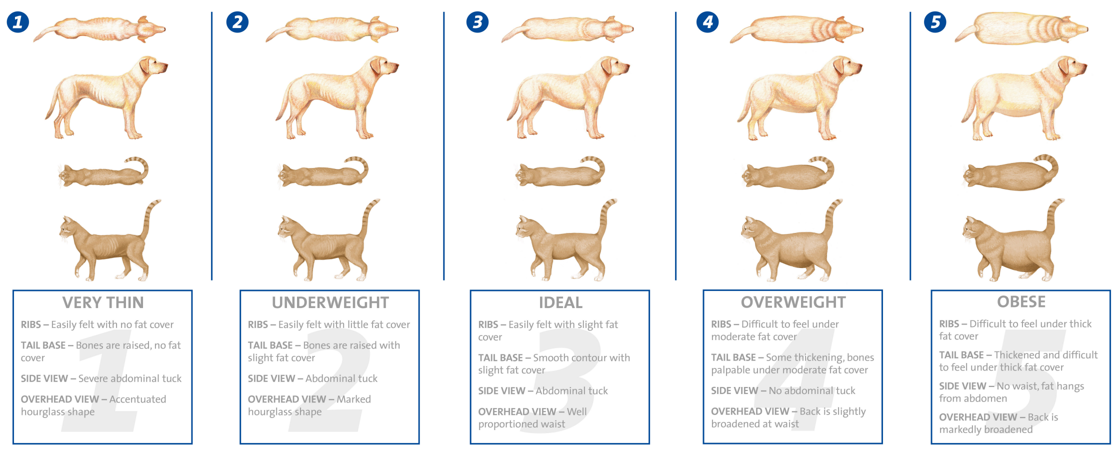


Figure 1.2: An example of a Body Condition Score [10]

In this scheme, levels can be distinguished with characteristics such as non prominent waist when viewed from above or ribs hardly palpable which and needless to say, those represent an evaluation prone to error, since most owners do not have the proper knowledge to evaluate his or her pet just by visual observation, leading to erroneous scores that can have

serious consequences to the animal's health.
It is possible to identify a need in the daily life of every pet owner that wants to monitor closely his or her pet health, more specifically its ideal weight condition. Owners may want their pet to lose weight and in this scenario, having real time feedback on weight loss caused by physical activity is essential, or they may want to increase the animals' weight and in this case understand if, for instance, the animal is getting an appropriate diet, or even in maintenance scenarios where keeping the animal's weight is of the owner's best interest.

When scenarios like these are considered, the creation of a cheap, easy to access and use system with a capability of giving real time feedback on the animal's activity level, translating or quantifying it into a single number such as the widely accepted unit of energy expenditure, calories, is a huge improvement in the management of a companion animal's weight. With this kind of information a pet owner can keep track of how much exercise his or her pet has been doing, it can improve the bond between these two parts making the life of the owner somehow more joyful and in some cases translate into more physically active habbits of the owner, ending up to not only managing the animal weight but also its own. An important side note to keep in mind is that even with such a system, no device should entirely replace the job of a veterinarian aware of the obese animal's evolution.

1.3.2 Pet Animals

This project is focusing its scope on canine pets and the main reason for that choice is the fact that with an attempt to encompass all the other companion animals, especially cats, not only changes in the algorithm parameters would have to be addressed but also behaviour, physical and diet changes would have to be considered.

Concerning physicality, cats have an instinctual desire to stalk quietly with lean muscles meant for leaping and at a very fast speed. Therefore their bodies are built for agility, rather than brute force or endurance. These traits contrast with the ones from a dog, an animal that relied on endurance to outrun his prey by running during large amounts of time.

Concerning the diet, cats rely on high protein meals to keep peak performance as well as eating with more frequency but smaller quantities in each meal, contrasting with dogs, almost omnivores that in serious survival situations may even be sustained by an entire plant-based diet [11]. They eat more infrequently but in bigger portions, taking slowly digesting processes and using the accumulated fat as energy very efficiently.

In short, cats are animals with very different activity patterns that would be hard to distinguish when compared to the ones of a dog.

1.4 Previous Work

This work takes advantage of an already initiated project done by the author on a summer internship at Findster Technologies, based on the adaptation of an already existing system, with the main goal of processing raw motion data from an embedded accelerometer in a device called *pet module*, worn by the companion animal. This data is sent through "MAZE", a proprietary protocol created by Findster Technologies, to another device, the *guardian module*, usually held in proximity to the smartphone by the pet owner. The *guardian module* finally communicates through bluetooth (low energy or BLE) with a smartphone transmitting only a number in the range of 1 to 3, corresponding this number to an activity level, as shown in figure 1.3.

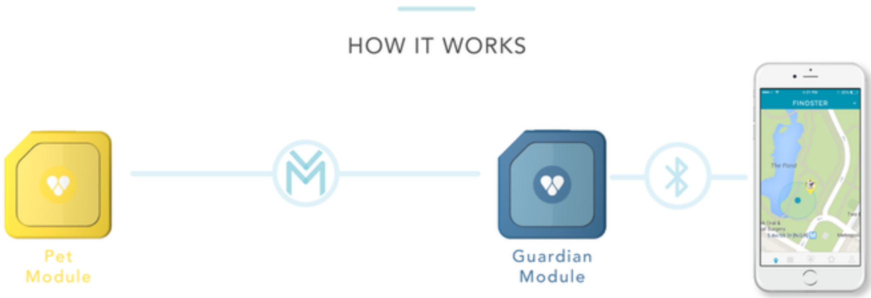


Figure 1.3: Modules used on internship's project

The integer values that are transmitted were, as stated, reflecting the activity level, where the number 1 referent to a completely stopped state, whereas the number 2 refered a moderate state of activity held by the animal, normally walking and number 3 meant that the animal was in a high activity level, generally described by activities such as running, jumping, etc. The activity level was refreshed every 5 seconds and all the processing steps of the data occurred inside the *pet module*, with a microcontroller with limited resources. The activity level was calculated based on the establishment of dynamic levels capable of separating three zones, each correspondening to one of three levels. During one time window, the acceleration values were summed and fit into one of those intervals for level estimation. The levels were constantly adaptated to give a degree of convergence to the system, that over time would reach an optimum state, translated into an optimum classification of activity state. This adaptation would occur every time an acceleration vector value over a time window surpassed the maximum registered activity. If that was the case, the levels would change their values in an attempt to better categorize activity. All these processing steps were based on simple logical operations, tipically simple *if then* blocks, easily perceived and understood. If the acceleremoter values were above the first level and under the second level, activity was classified as moderate, or if the maximum acceleration value in an instant was higher than the established maximum, the maximum variable would be replaced and the levels would readjust. Even with this technique, an estimation of the best initial value was needed and the assessment or discovery of such value was one of the biggest challenges of the problem at hand. Many factors such as animal's age, gender, breed, size or weight influence directly the oscillating capabilities of the animal, with a consequent importance on physical state (since it is measured with an accelerometer) and so the levels do also mirror that influence. With all these factors considered and based on a small group of samples, three base values were determined for the starting point of each level, with the intention to encompass all variables. However, a significant improvement was needed because such values do not produce acceptable results across dogs with different sizes, resulting in misleading classifications.

1.5 Main Team

This project is born through a joint collaboration between Findster Technologies and the University of Coimbra (UC), more specifically LIBPhys-UC.

1.5.1 Findster Technologies

Findster Technologies is a startup company, founded in March of 2015 [12], born from a necessity of keeping track of a companion animal that was always escaping its own area and getting lost in the process. Its commercialized product is a GPS pet tracker composed of three parts, a *pet module*, a *guardian module*, and a smartphone application, similar to the system described earlier. In *Findster Duo* the *pet module* is coupled to the companion animal and its position is constantly recorded and the data sent through radio frequency (RF) to the *guardian module* that communicates by bluetooth with the smartphone application able to display a map and the current position of the *pet module*, as shown in figure 1.4. With this system the user is capable of, through the application, create fences that delimitate the safe zone of its pet and if that zone is trespassed be warned by a generated alert, steering the owner towards its animal.

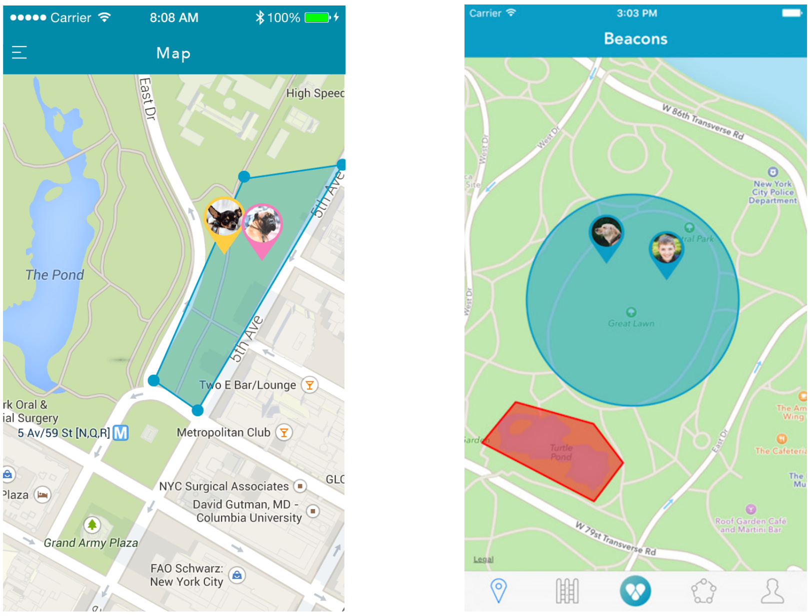


Figure 1.4: Application screenshots with Findsters tracked within a customized fenced

The major differences between *Findster Duo* and already available systems are the removal of monthly fees since there is no Sim card in any of the modules, freeing the user of costs related to the number of times the system communicates, the real time communication and the fact that it is a standalone system, meaning that it can be used in places with no signal coverage.

1.5.2 LIBPhys-UC

The Laboratory for Instrumentation, Biomedical Engineering and Radiation Physics (LIBPhys http://libphys.fis.uc.pt) is a joint research center located on the Physics Department of the University of Coimbra. Its activities are focused on radiation physics, biomedical engineering, fundamental parameters and analytical methods, it is composed by researchers and fellows from 3 institutions (University of Coimbra, NOVA University of Lisbon and the Research center for Oral and Biomedical sciences from University of Coimbra). In the Biomedical Engineering field its activities in Coimbra's University have been focused on Hemodynamic Instrumentation and Analysis, Biometric and Clinical signal processing and Malaria instrumentation for RDT (Rapid Diagnostics Tests)

1.6 The work

The goal of this work is to implement a robust and functional activity tracker, that can be added to the already existing system.

1.6.1 Key Contributions

The key contributions of this work are:

- A classifier that is able to assign 1 out of 5 activity levels in real time
- A pedometer capable of counting the number of steps given by the dog
- An equation that relates energy expenditure and acceleration, used to calculate spent calories in real time

1.6.2 Deliverables

This work also has a smartphone application developed for Android operating system, capable of displaying the data of the previous sub-section.

1.7 Thesis Outline

This thesis is divided in 8 chapters.

Chapter 1 gives an initial context of the addressed problem and identifies flaws on current solutions while presenting the motivation to create a better approach. It also introduces the involved team members.

Chapter 2 explains the meaning of some frequently used terms in activity assessment while exposing used methods to calculate energy expenditure in humans. It gives an overview on the existing energy expenditure methods with focus on error, cost and ease of use. It also gives general guidelines regarding weight management focusing on the animal's diet and shows the system's potential for purposes other than companion animals. Finally it describes relevant features that can be extracted from accelerometer signals alongside classification methods that can use these features to describe different types of activity.

Chapter 3 gives an overview on state of the art both in commercial and experimental solutions.

Chapter 4 gives a detailed and categorized list of the system's requirements and architecture.

Chapter 5 explores in greater detail the pedometer's algorithm design, explaining used technologies, data acquisition techniques and briefly introducing the processing methods applied on obtained signals.

Chapter 6 explores in greater detail the creation of a relation between acceleration and caloric expenditure. It introduces the way measurements are made, their difficulties as well as the processing techniques applied *a posteriori*.

Chapter 7 recalls the used steps in the two previous chapters and illustrates the obtained results with some performance metrics.

Chapter 8 presents the created system's main advantages, its limitations and what topics to focus if future work is to be considered.

Chapter 2

Background: Physiology, Behaviour and Analysis

2.1 Activity Assessment

Once identified the work's objectives it is relevant to explore the current ways of qualifying and quantifying physical activity.

In the research field, the widely used methods of measuring activity take advantage of a metric called energy expenditure (EE) or total daily energy expenditure (TDEE), both represented in calories, an energy unit that gives the needed energy to raise the temperature of one gram of water by one degree Celsius at a pressure of one atmosphere [13]. Even though these terms are often used to describe the same, there is a difference between them, since energy expenditure can be equal to basal energy expenditure, BEE (synonymous with basal metabolic rate BMR) if the amount of energy or calories necessary to carry out basic functions like breathing, digesting or absorbing food is considered, or it can be equal to total daily energy expenditure if the total number of calories burnt per day, including other activities such as physical exercise is accounted [14]. This means, for example, that a sedentary individual will have his BEE proximate to its TDEE. Other equally important terms to define are Resting Energy Requirements (RER), Maintenance Energy Requirements (MER) and Metabolic Equivalent Task (MET).

RER, similarly to BEE, does not account any type of physical activity but is measured with less restricted conditions, given that to obtain BEE, the subject must be in a darkened room upon waking after 8 hours of sleep and 12 hours of fasting. Despite the proposal of many formulas for the calculation of this parameter, the standard in veterinary medicine is the predicitive equation descendent by Kleiber's law (equation 2.1), where BW stands as the body weight in kilograms.

$$\text{RER(kcal/day)} = 70 \times \text{BW}^{0.75} \tag{2.1}$$

$$\text{RER(kcal/day)} = 30 \times \text{BW} + 70 \tag{2.2}$$

Max Kleiber, a swiss agricultural biologist, found that in the vast majority of animals, metabolic rates scaled to the 0.75 power of their mass and this relation was valid for large ranges of body weight from cases such as a 10 gram mouse to a 1 ton whale [15]. The fact that it is a simple equation, made it the most common method to calculate RER, even though it has some limitations, namely in injured or hospitalised dogs, where the power coefficient may be slightly different. The first equation shows a more generalized approach that can be used, theoretically, in any animal and the second one in subjects with lower weights, however in some cases it may overestimate the caloric needs [16].

In humans several prediction equations have been tested, but the one with narrowest error range is the Mifflin-St Jeor equation (equation 2.3) [17] which takes more inputs providing an error of less than 10%

$$\text{RER(kcal/day)} = (9.99 \times \text{BW(Kg)}) + (6.25 \times \text{Height(cm)}) - (4.92 \times \text{Age(years)}) + 5 \quad (2.3)$$

In contrast with Kleiber's law, Mifflin-St Jeor equation has different coefficients for male and female individuals, bwhere the ones above are for male humans.

Once an individual starts to be active, it also starts to burn more calories (same as saying spending more energy) than the ones calculated in equation 2.1 therefore demanding a bigger calorie intake. The calculus of MER, made in different ways accounts for this increase in caloric demand, and in this case there is not a formula that can be referenced as the golden standard one, so relevant examples are shown.

One hypothesis is that MER can be calculated by multiplying RER with a factor, representative of a growth stage by the animal (early age), its activity state and intensity, if the animal is neutered or not (neutered dogs have smaller energy needs [18]), etc.

$$\text{MER} = \text{a} \times \text{RER} \quad (2.4)$$

Where "a" is the multiplying factor. Needless to say that if a robust model is to be created, a simple factor is not able to reproduce the many variables inherent to different situations in each activity measurement.

MET is a measure that expresses the energy cost of physical activities, defined as the rate of energy consumption during a specific physical activity to a reference metabolic rate.
In mathematical terms:

$$1\text{MET} = 1\frac{\text{kcal}}{\text{kg} \times \text{h}} = 4.184\frac{\text{kJ}}{\text{kg} \times \text{h}} = 1.162\frac{\text{W}}{\text{Kg}} \quad (2.5)$$

Most of the times, MET data for human activity is in table format, similarly to table 2.1 where several activities are considered and its energy cost is expressed in MET units to calcululate the energy expenditure.

Table 2.1: MET table translating specific physical activities into energy cost [19]

		Corrected METs - Female		Corrected METs - Male	
Activity (code)	**2011 Compendium METs**	Normal Weight 60 Kg, 168 cm, 35 yrs	Overweight 77 Kg, 168 cm, 55 yrs	Normal Weight 70 Kg, 178 cm, 35 yrs	Overweight 91 Kg, 178 cm, 55 yrs
Rope jumping (15550)	**12,3**	13,5	16,5	12,9	15,4
Running, 6 mph (12050)	**9,8**	10,7	13,1	10,3	12,3
Bicycling, general (01015)	**7,5**	8,2	10,0	7,9	9,4
Pushing stroller (17100)	**4,0**	4,4	5,4	4,2	5,0
Calisthenics (02030)	**3,5**	3,8	4,7	3,7	4,4
Shopping (05065)	**2,3**	2,5	3,1	2,4	2,9
Watching TV (07020)	**1,3**	1,4	1,7	1,4	1,6
Total MET-min	**1,221**	1,335	1,635	1,294	1,530

In order to accurately infer total daily energy expenditure, three main factors should be considered, the basal metabolic rate which is, as said before, the minimal rate of EE per unit of time; thermic food effect, meaning the increase in EE associated with digestion, absorption and storage of food representative of generally, 10% of the BMR and activity thermogenesis or the process of heat production in an organism. This term is divided in two main branches, the exercise-associated thermogenesis (EAT), caused by the engagement in physical activities and the non-exercise activity thermogenesis (NEAT), caused by other factors such as, for instance, body temperature maintenance, spontaneous muscle contraction, posture maintenance and other daily activities. Once again, a sedentary individual that executes no exercise whatsoever will see his or her EAT be equal to zero.

Beyond these three main pillars of energy expenditure, others may be considered in some cases, like medication, emotional state or altered temperature, meaning that EE is highly variable even in the same individual.

2.1.1 Measurement Methods

Following an already established division of EE measurement methods, these fit into one of three blocks; direct calorimetry that measures directly, as the name illustrates, heat loss from the subject; indirect calorimetry, where the consumption of oxygen and/or carbon dioxide production is measured and translated into EE and non-calorimetric techniques that try to find a relation between physiological measurements and EE, attempting to extrapolate energy expenditure from observations [20]. These techniques are the ones with emphasis given throughout this work.

2.1.1.1 Direct calorimetry

- **Isothermal systems** - An isothermal calorimeter is composed of a chamber with an insulating material maintained at a constant temperature with circulating fluid. The temperature gradient across the defined material is proportional to the heat loss taken place in the subject and the response time could be under 5 minutes and the error 1%.

- **Heat sink or adiabatic systems** - The principle used by this sort of systems is the base of equipments such as calorimeter suits, where a chamber from which heat lost by the subject is extracted by a liquid-cooled heat exchanger and the rate of this extraction is regulated so that the temperatures in both sides of the chamber's walls are equal. Response time is around 30 minutes and an error less than 3% is verified.

- **Connection systems** - These calorimeters also have an insulated chamber but are ventilated with air at a known rate. This flow rate allows the calculation of the energy expenditure from the increase in temperature of the ventilating air (caused by the subject's heat loss). The response time is 20 minutes and the error 1 to 2%.

These kind of instruments are extremely expensive to build and to run, while also needing an expert during the whole process.

2.1.1.2 Indirect calorimetry

- **Total collection systems** - In these systems, the subject expires air to one component that can be an airtight rigid structure suspended over water whose height increases and the air composition changes as the subject expires, characteristic of rigid total collection systems, or a portable flexible bag connected through tubing to a valve that enables the entrance of expired air at first and then is turned to let the air exit to be analysed, characteristic of flexible total collection systems. Both these systems represent an error lower than 3% but the measurement time is quite long (2 hours in some cases).

- **Open circuit indirect calorimeter systems** - In these circuits, the subject not only expires to a component but also inspires air from a structure, with an analysis on expired gases. The expired air is taken out with a pump, mixed with atmospheric air, dried and studied for oxygen and carbon dioxide concentrations. Oxygen and carbon dioxide are measured by the use of paramagnetic and infrared respectively or using a mass spectrometer. The test time can vary from 30 seconds to 30 minutes depending on the composition of the system.

- **Respiratory chambers** - In this case, the subject is placed inside a chamber with a known air volume and the oxygen comsumption and carbon dioxide production are measured from the gas concentration changes inside the chamber. These methods involve an error of 2%, but are rarely used nowadays [20].

- **Closed circuit systems** - A good example of this kind of system is a spirometer, consisting of a bell containing oxygen, inspired by the subject. The bell is suspended over water and its height decreases in a proportional rate of the oxygen consumption.

2.1.1.3 Non-calorimetric techniques

- **Doubly labelled water** - This technique makes use of non-radioactive isotopes that ”mark” both hydrogen and oxygen of water that is inside the subject, which the body eliminates and this rate is used to estimate carbon dioxide production as well as EE.

$$CO_2 + H_2O \longleftrightarrow H_2CO_3 \tag{2.6}$$

Oxygen is ”marked” with O^{18}, this isotope will distribute through body water, circulating in H_2CO_3 and in the expired CO_2. As the carbon dioxide is expired and body water is lost in urine, respiration and perspiration, the concentration of the label decreases. The same principle is applied to hydrogen, which is labelled with a tracer whose concentration decreases over time as body water is lost. If both elements are tagged, differences in the elimination rate will allow the establishment of the CO_2 elimination. Samples of urine, saliva and blood are analyzed through mass spectroscopy and the doubly labelled water is administered orally. Test time period can be 7 to 21 days and an errod of 6 to 8% is associated.

2.1.1.4 Subjective methods

- **Activity recall** - When non-specific information is considered, questionnaires as well as interviews can be used to assess activity, although substantial error is introduced thanks to inadequate data recording or innacurate recall.

- **Activity logs/factorial method** - Used many times to estimate activity thermogenesis, physical activities are logged during a time period and the energy of each activity is obtained through the use of tables similar to the one in 2.1 that is multiplied by the hours spent in each activity. The sum of all energies and the combination with information such as the basal metabolic rate allows to estimate total daily energy expenditure (TDEE). Once again, sources of error exist and the main ones are related to innacurate recording of activities and determinations of their energy cost, the inexistent guarantee of precise and accurate activity tables. Futhermore, dog owners may be illiterate, report incompletely their activity patterns during the assessment period. Even though training of dog owners may reduce significantly this error, that strategy is cumbersome and time consuming.

2.1.1.5 Objective methods

- **Integrated eletromyography** - In this system, cumulative eletrical muscle activity from muscle fibres is measured over a defined period but the relationship between different groups of fibres and EE varies making it necessary to measure multiple muscle groups at the same time.

- **Heart rate monitoring** - This technique takes advantage of the relation between energy expenditure and heart rate, even though their relation may not be linear, since the ejected volume can change with different heart rates and other factors like posture, emotion, between many others. Furthermore, there is a considerable inter and intra individual variance meaning that the precision of these systems is not as good as some of the previous. An option to improve it is to use an equations for each subject, representing a very specific approach with errors, at best, of 3% to 20%.

- **Kinematic measurements** - Finally, subject movement's can be analysed in order to quantify energy expenditure, normally used when describing spontaneous physical activity. These techniques focus on the use of instruments such as pedometers and accelerometers of different sophistication. Pedometers are devices that basically count steps, detecting the displacement of a subject with each stride, but since they are not able of quantifying total body displacement, the prediction of activity thermogenesis is poor [20]. On the other hand, accelerometers are able to detect body displacement with varying degrees of sensitivity and in different directions, where higher accuracy obtained in three axial accelerometers (rather than one axis).

2.1.2 Canine Pets

Over the course of this project the assessment of activity level is going to be focused on companion animals more specifically in dogs. It is expected that the developed methods are robust enough to incorporate inter subject variability caused by changes in race, gender, age and weight.

2.1.2.1 Gaits

The human definition of gait is considered as a manner of walking, stepping or running and in terms of a dog it is a quality of movement. Canine's gait is characteristic of each breed and may be decomposed in different styles, where the two main groups are symmetric and asymmetric. With symmetric gaits such as walk, trot and pace, one side of the animal's body represented by two limbs repeats the motion of the limbs on the opposite side with the intervals between paw falls almost evenly spaced. When considering canine gait, one full cycle is often referred as the stride. Contact time between the paw and the ground is directly proportional to leg length, meaning that long legged dogs will have a longer contact time than short legged ones. There is a great deal of variation in the gaiting pattern according to size and among individual dogs, but generally, small dogs show the greatest variation in their gait [21]. Once an owner is capable of ”recognizing gaits and anticipating patterns as the animal is moving, it is possible to identify gait abnormalities, distinguish abnormal from normal gait variations and start to identify sources of gait disorders” [22].

- Walk - The least tiring gait, walk has four beats in which each limb steps sequentially. It can be a normal walk, where body weight is supported by two and three limbs alternatively, neck and head are lowered during forelimb swing and raised during forelimb support phase. Stride length is such that the back paw approximately overlaps the place occupied before by the front fore paw. It can also be a power walk, when the dog is pulling a load, with shorter and slower steps, and the support by three limbs predominant where the head may be lowered and fore limbs contribute significantly to propulsion.

- Amble - Essentially an accelerated walk where limbs move faster, the support is made alternatively with one and two limbs.

- Trot - Gait mostly known in horses, trot is a two beat gait characterized by alternating diagonals supporting the weight, *i.e.* right fore and left back paw are in contact with the ground followed by the left fore and right back paw. If a running trot is considered, there is a suspension phase between both diagonals where none of the paws is touching the ground, on the other hand if it is a slow trot, the suspension phase is inexistent.

- Pace - Also a two beat gait, pace has two limbs to support the dog’s weight but this time, both limbs (fore and back ones) are on the same side, *i.e.*, first the right fore and back limbs touch the ground followed by the left fore and back limbs. A short period of suspension happens between the size change, faster in what is called a faster pace and slower in a slow pace.

  Although a good equilibrium and less vertical oscilation are achieved, trot shows to be more stable [22].

- Canter - This faster gait is characterized by a sequence of two support limbs followed by one only. The first phase is made with a diagonal support and the next one is supported by a suspension characteristic of a fast movement. It is an assymetrical gait and therefore is used by dogs when changing direction or looping and it is suited for rough ground or where footing is uncertain.

- Gallop - Characteristic of the fastest movement of the dog, this gait is expressed by one limb of support or none, in a suspension phase. The forelimb supports weight just before suspension and the hindlimb that impacts the ground first is extended forward. It is therefore more prone to injury because of the shock involved in catching the falling body weight (expressed in the schematics of figure 2.1). It can be subdivided in transverse and rotatory gallop where the last one the fastest and most fatiguing of all gaits. Using this gait the greyhound can achieve speeds of 60 Km/h.

  Trunk flexion is also involved in the locomotion, where a flexible trunk is capable of achieving bigger strides during the extension and contraction phases of suspension.

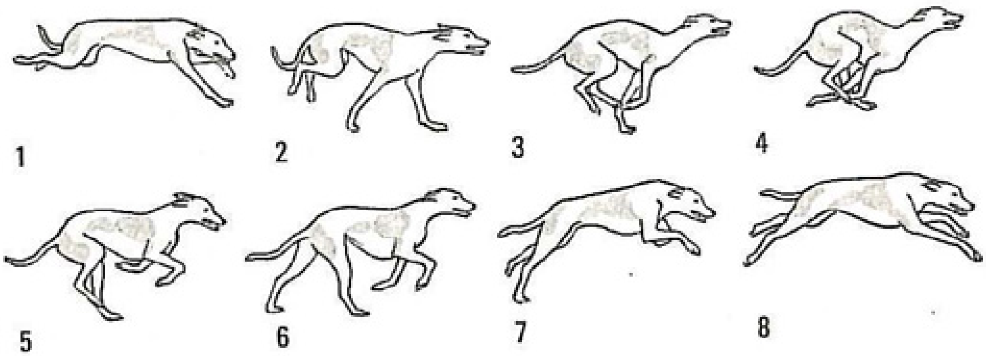


Figure 2.1: Greyhound schematics of gallop gait with trunk flexion in 4 and extension 8 [23]

2.1.2.2 Monitoring Gaits

Considering all previously described gaits, it is possible to infer that a proper, real time distinction between gaits like Pace and Canter is extremely difficult, reason why there are some key characteristics to help in such classification. With the intention to facilitate this distinction, a five category division is proposed in this project.

- **Stop** - Representative of a recumbent dog with no activity at all.
- **Walk** - When the dog is moving slowly, supported by two or three paws, tipical of little energy expenditure.
- **Trot** - Fast walking where the dog has always a front limb and a hind limb on suspension. This state encompasses the characteristics of trot and pace, described previously.
- **Run** - Equivalent to canter, this state is initiated by a running state by the dog.
- **Intense run** - Characteristic of an intense run with visible effort by the animal, this state encompasses the transverse and rotatory gallop also described earlier.

Where these gaits are the ones to be classified by the system.

2.2 Weight Management

As it was said, excessive weight can reduce life quality as well as longevity, both in animals and in humans. Once a dog is overweight, the question is no longer "if" the dog or cat is going to develop a secondary condition to the excessive weight, but "when". The most common disorders are skin, heart and respiratory related, renal dysfunction, orthopedic like osteoarthritis, metabolic and endocrine disorders like diabetes and many forms of cancer, especially intra abdominal [16, 24]. Moreover, overweight dogs tend to physically interact less with others while less energetic and playful, and this inactivity may hide illnesses that in other cases are easy to identify, therefore it is of major importance to raise awareness on negative health consequences of excessive weight since more than 30% of dog owners have misconceptions about ideal weight, considering their dogs to be at an acceptable weight, when they are not [16]. If a dog is overweight, obese or underweight, a weight management plan is often the best option, but no plan can be started without the assistance of veterinarian healthcare teams, due to the fact that medical conditions like hypothyroidism, hyperadrenocorticism or cushing's disease (an increase in cortisone due to a benign tumor of the pituitary gland) may be the main cause of excessive weight. If these conditions are not the case, there are some guidelines provided by associations such as AAHA (American Animal Hospital Association), or APOP (Association for Pet Obesity Prevention), composed by groups of veterinarians that give easy and practical tips on how to create an individualized plan for the animal to return to its ideal weight. Although this plan must encompass physical activity (focused part in this project), it must also consider the other pillar of weight maintenance in animals and humans, dietary regime, often translated to caloric restriction and diet selection.

2.2.1 Caloric Intake

For safe weight loss in dogs, it is recommended a 3 to 5% body weight loss per month and the way to do it is feeding the animal with 80% to 100% of its resting energy requirements (described in equation 2.1) and in cases that it fails to respond, with the veterinarians approval, the total number of calories may be reduced [24]. To determine if a plan is responding or not, a close monitoring must be held, with a monthly weighing and a weight loss bigger than one pound. The majority of dogs achieve their ideal weight within six to eight months, so a longer process and a monthly weight loss inferior to one pound are good indicators that another approach might be of good use. Also the adaptation time when changing a diet is very important and it is recommended that the switch is made gradually with a quarter of the total reduction in one week, half in the second, three quarters in the third week and a total change after a month.

2.2.2 Food Composition

Once the quantity is established it is also needed to select what type of food is going to be given to the animal. A lot of options are available nowadays, from dry or canned food, therapeutic food, even though these are only available through a veterinary office or with a prescription since their main goal is to modulate a disease or condition. When portion size is reduced, alongside calories, essential nutrients decrease and since the animal is intended to practice more physical activity, the protein content of the diet must be known so that enough energy is provided to the animal, meaning that a high protein content may preserve muscle mass during weight loss and furthermore may improve satiety, defined as the feeling of fullness and disappearance of appetite after a meal [25]. However, diet protein content is not a linear issue since older pets may have more difficulty in digesting proteins requiring closer monitoring.
There are also other solutions related to medication, with supplements and other pharmaceutical solutions for the management of obesity in dogs, but that approach is out of this project's scope.

2.2.3 Physical Activity

Caloric expenditures for different forms of exercise in pets are largely undocumented [16], meaning that the existing solutions are based on practical guidelines such as, starting with a five minute walk three times per day and increasing gradually until a total of 30 to 45 minutes (per day). These sort of guidelines, applicable to obese dogs not possessing orthopedic diseases are broad and try to encompass all obesity cases, while in some of them may understimate the animal need for energy spending.

2.2.4 Follow-up Procedure

As important as changing a dog's mass to its ideal weight, is maintaining it once the goal is achieved. To calculate the percentage of loss weight the formula is:

$$\text{Weight Loss (per week)} = \frac{\text{initial weight} - \text{actual weight}}{\text{initial weight}} \times 100 \qquad (2.7)$$

where the ideal values are between 1 to 2%/week for dogs (and 0,5 to 2%/week in cats) except for growing dogs with age under one year. If the plan is not efficient a reduction on calorie intake may take place but a lower limit (usually 60% of RER of ideal body weight) must be established so that nutritional deficiencies do not take place and when the ideal weight is achieved the best approach is to keep the same amount of calories within the diet and only if the dog keeps loosing weight, than an increase of 10% can be made.

2.3 Farm and Wild Animal Activity Tracking

Even though the focus is on canine pets, the potential of a system like the one described in this project goes beyond companion animals and activity level monitoring. These systems show themselves as an extremely useful measuring tool in difficult circumstances for human observation, such as aquatic [26, 27] and aerial environments. Even in terrestrial species, these systems may be used to record behaviour automatically in remote locations without human presence. A few examples of already tested and implemented systems make use of 3-axis accelerometers to classify not only physical activities but behaviour patterns, useful in other topics such as defficient walking, physical distress, etc. Perhaps it is a relevant future work to try and implement an even more robust systems that can be applied throughout a bigger group of species and still be able to identify patterns.

2.3.1 Farm Animals

On farm animal behaviour assessment is difficult, especially in large herds, due to the time constraints and lack of labour [28]. The already in use systems are suited to only classify one or two behaviour patterns, with a more complex system, capable of automatically distinguish several different patterns needed to assess better animal health and welfare. Other than a health indicator, livestock behaviour can serve as a tool for improved management to enhance animal performance in reproduction or lactaction and the study of specific activities like graze may enhance the human understanding of animals' use of vegetation [29]. One approach in this field is the use of accelerometer data and support vector machines as a classifier, to distinguigh between standing, lying, ruminating, feeding, normal walking, lame walking, lying down and standing up in cows in a farm environment [28]. Another example is an automated, simple behaviour analysis system for goats' activities at pasture where resting, walking and eating were classified and eating still subdivided in grazing and browsing, by the difference in head posture [29].

2.3.2 Wild Animals

2.3.2.1 Aquatic environment

Analysis of behaviour patterns in underwater environments is limited since assessment of time spent swimming and swimming speed by free ranging fish are difficult to obtain. Some alternatives rely on direct underwater observation by scuba divers and/or video cameras, or water speed sensing transmitters, but these remain limited to situations like nocturnal fish, water currents that influence speed sensors, not considering the overestimation of swimming speed and energy expenditure by these kind of sensors. Therefore a good approach for this problematic is, once again, the use of systems with accelerometer sensors capable of giving an insight on aquatic animals' such as penguins, sea turtles [30], migrating fish, between others [27, 26]. Other than activity states like resting, swimming, gliding and feeding, also swimming speed, diving angle and diving depth can be obtained from these systems, helping to keep a close monitoring of, for instance, endangered species.

2.3.2.2 Terrestrial environment

An already tested system was implemented on elephants, both in wild and captive environments to classify four behaviour patterns such as feeding, walking, bathing and swaying [31]. This system using accelerometer alongside GPS data is able to measure walking distance, identify activity patterns and trigger wireless transmissions, if there are signs of physical distress, unusually high or low activity or limping, that are sent to responsible personnel able to properly deal with the situation. The results show not only a reduction of the lag time between an event and aid delivery, but also can avoid worse situations such as illegal hunting. Also in captive environments, the system can also make use of the accelerometer data to infer physical activity levels (like the ones assessed in this project), time spent during those activities in order to guarantee a proper health management.

2.4 Activity Data Analysis

2.4.1 Parameter Calculations

Other than the obvious values obtained by the sensors used in the project, more specifically the 3-axis accelerometer that illustrates kinetic movement by the animal, there is a much bigger quantity of information that can be withdrawn, even if not directly. This kind of information of which the average acceleration in one of the three axis is an example, are often called as features, since they are calculated values that represent information of a time period in a single number, contrasting with the raw values from the sensor that simply describe oscillation in a very short instant. With that said it is important to gather features that are used in similar systems that make use of accelerometers with the purpose of measuring motion patterns. Different group divisions are made in projects similar to this one, but in the majority of cases, both time and frequency domains are used in order to generate features.

2.4.1.1 Time domain

A time window with a defined length is selected and characteristics are obtained:

- Movement intensity - also refered as signal magnitude vector, this value measures the instantaneous intensity of movements at instant t, as well as the gravity acceleration [32, 33]:

$$MI(t) = \sqrt{x(t)^2 + y(t)^2 + z(t)^2} \tag{2.8}$$

  where $x(t)$, $y(t)$ and $z(t)$ stand for the body acceleration in x, y and z axis respectively.

- Mean (MA) acceleration - this value can be calculated for each individual axis or can be calculated for all three [33, 34, 35, 36, 37]:

$$\mathrm{MA} = \frac{1}{T}\sum_{t=1}^{T}\mathrm{MI}(t) \tag{2.9}$$

- Variance and standard deviation (SD) of acceleration - a measure of the variability of the points in a time window [36]

$$\mathrm{Variance}(X) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \quad , \quad \mathrm{SD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{2.10}$$

- Mean absolute deviation (MAD) - the average of the differences of each point to the mean acceleration
- Mean value of local maxima - used to distinguish stronger from weaker activities, it is calculated by the average of a defined number of activity peaks [38]
- Skewness - third measure of the probability distribution's assymetry [34, 35]
- Kurtosis - another metric to describe the probability distribution of the variable depicting the tail's form of this distribution [34, 35]
- Time between peaks - the distance between activity peaks assumes an earlier definition of these that can be done in different ways, a proposed one in [36] is to obtain the maximum value, correspondent to the biggest peak in the time window, establish a threshold based on a percentage of it and the local maximums of the acceleration curve above the threshold are considered as activity peaks.
- Step count - in order to obtain this feature, special care had to be taken once the device placed on the neck collar was not in a fixed position, resulting in a constant change of the gravity acceleration's direction [38]
- Area between mean of local maxima and signal - for this feature, the mean value of local maxima was calculated and the area between that value and the acceleration curve is obtained [38]
- Crossing of mean value - used to separate less dynamic from more dynamic activities this feature is the count of times the acceleration curve crosses the mean value of acceleration [38]
- Signal magnitude area (SMA) - defined by the area below the acceleration curve, this feature is used as an indirect estimation of energy expenditure [32, 33]

$$\mathrm{SMA} = \frac{1}{t}\left(\int_0^t |x(t)|dt + \int_0^t |y(t)|dt + \int_0^t |z(t)|dt\right) \tag{2.11}$$

where $x(t)$, $y(t)$ and $z(t)$ are the acceleration components. This feature can also be considered as a measure of the velocity in a time segment since it integrates the acceleration. In some cases the velocity is calculated but only in one direction [33]

2.4.1.2 Frequency domains

- Energy - since in the time domain this feature obeys to the formula

$$E_s = \int_{-\infty}^{+\infty} |x(t)|^2 dt$$

where in the frequency domain this feature is calculated by:

$$E_s(f) = |X(f)|^2 \tag{2.12}$$

where X(f) is the Fourier's transform of x(t), used in [35, 37]

- Entropy - a complex measure of the signal's disorder ou uncertainty. A possible way of calculating it is through

$$H(X) = -\sum_{i=1}^{n} P(x_i)\log_b P(x_i) \tag{2.13}$$

where b is the base of the used logarithm (usually the value 2 is chosen), and $P(x_i)$ the probability that $X = x_i$

- Spectral centroid - after applying a fast fourier's transform (FFT) to the signal, a curve in the spectrum is obtained. The centroid of the object defined by the area under this curve is used as a feature [33]. Its purpose has to do with the frequency that the centroid corresponds, for instance, a centroid placed in high frequency levels associated with vigorous movement

- Spectral spread - a measure of how spread is the spectrum obtained after, for example, a FFT is applied to the time window [33] calculated by the diference of the obtained spectrum and a horizontal spectrum with height equal to the signal's maximum

- Frequency peak - this feature is the frequency obtained by the highest point in the time window's spectrum, it stands as the x-coordinate

- Power of frequency peak - the same as the previous feature, but this time the y-coordinate is taken [34, 39]

- Signal power in different frequency bands - this approach considers the spectrum, divides it into several zones, accordingly to established frequency values and calculates the power, *i.e.* the area between the spectrum's curve in that zone [34]. A note worth mentioning is the fact that this feature will not return a single value but instead a group of values, and this calculus can be done with or without normalizing the values to the window size [33]

- Cepstral coefficients - a different feature used in [37] defined by FFT(log(FFT(x))) where x is the acceleration data in a time window

- Spectral energy entropy - the energy and entropy calculations (described earlier) are done, but this time on the signal's frequency spectrum.

2.4.1.3 Other

Other features do not possess characteristics explicitly from neither of the previous domains while still be worth mentioning.

- Day of the week, time of day and whether the day is weekend or not - a group of discrete values representing in the first case either morning, afternoon, evening, night or late night. [40]
- Activity in the previous and posterior time window
- Correlation coefficients - coefficients obtained by the correlation matrix of an axis pair, xy, xz and/or yz [37]
- Eigenvalues of dominant directions - where the dominant directions the ones with more oscillation [37]

2.4.2 Classification Methods

After the extraction of a given number of features, it is crucial to give them meaning. To do so, different types of algorithms are tested in an attempt to find repeated patterns on the data.

2.4.2.1 Regression algorithms

Regression analysis is a statistical process, composed of algorithms that try to estimate relationships between a dependent variable and one or more independent variables. The dependent variable is changed while the others remain fixed and the estimation target is a function of the independent variables called the regression function, where the probability distribution is the variation of the dependent variable around this regression function. These techniques are used for prediction and to understand in the group of independent variables which ones are related to the dependent variable as well as exploring these relationships. They are often used to estimate continuous variables.

One commonly used example is the linear regression, that states a dependent variable (y_i in this case) is equal to a linear combination of parameters ($\beta_i, i = 1, ..., n$). In a simple linear regression when there is only independent variable (x_i)

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, ...n \tag{2.14}$$

where ϵ is the difference between the real value and given value by the model (error), $e_i = y_i - \hat{y}_i$.

In order to calculate the parameters, an objective function must be minimized and the most used one is the sum of squared errors (SSE)

$$SSE = \sum_{i=1}^{n} e_i^2$$

Considering the previous example of simple regressions the parameters can be calculated with the formulas:

$$\beta_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad \text{and} \quad \beta_0 = \bar{y} - \beta_1 \bar{x} \tag{2.15}$$

where $\hat{x}$ and $\hat{y}$ are the average of the x and y values respectively. Finally to calculate the model error, it is often used the mean square error (MSE).

$$\sigma_\epsilon^2 = \frac{SSE}{n-2} \tag{2.16}$$

where the denominator is the number of points subtracted by the number of parameters. However, most relations are not linear and contain information of more than one independent variable. Therefore a more general multiple regression model with p independent variables:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \epsilon_i \tag{2.17}$$

where $x_i j$ is the i^{th} observation of the j^{th} independent variable and the error is

$$\epsilon_i = ||y - \sum \beta_i x_i||^2 \tag{2.18}$$

and to calculate the parameters, in matrix notation:

$$y = X\beta <=> X^T y = (X^T X)\beta <=> (X^T X)^{-1} X^T y = (X^T X)^{-1}(X^T X)\beta \tag{2.19}$$

since $(X^T X)^{-1}(X^T X) = I$:

$$I\beta = (X^T X)^{-1} X^T y <=> \beta = X^+ y \tag{2.20}$$

where X^+ is the Moore-Penrose pseudoinverse.

2.4.2.2 Instance based algorithms

The models created from these type of algorithms compare new data with the already seen instances in training, stored in memory. It constructs hypothesis directly from the data instead of making a generalization like other methods, therefore the complexity of the model is intimately related to the complexity of the data, ans a small group of data usually is not the best option in this approach. Although, an advantage may be the model's capability to adapt with unseen data. These algorithms often store the training set and when a new instance is to be classified, distance metrics or similarities are calculated between the two elements to support a decision. That is precisely what occurs in the commonly known K-nearest neighbour (KNN) algorithm, a non-parametric method used both in classification and regression. In classification problems, when a new instance is introduced to the algorithm, it computes the distance to the K closest instances and a voting choice is applied where the most frequent class is chosen to classify the new instance. In regression problems, the instance value is computed by the average values of its K neighbours. To calculate the distance, different metrics are valid such as euclidean or manhattan distance.

2.4.2.3 Decision tree algorithms

These algorithms, with the construction of a tree diagram, are known for its use in classification since its major application is to classify instances in discrete values, even though some trees can return real values (an average of the criteria is calulated), the regression trees. In classification problems these algorithms have a greedy, divide and conquer approach where a tree is constructed top down, *i.e.*, every example is divided recursively based on the attribute that contains more information (dividing more clearly the data). In order to check which attribute is the best, a heuristic or statistic measure is chosen and where the most common is information gain. The main goal of a decision tree is to separate the data so that all instances in the last branch are of the same class (hence the name top down). The most common decision tree algorithms such as iterative dichotomiser ID3, or C4.5 use as heuristic the information gain:

$$\mathrm{Gain}(S, A) = \mathrm{Entropy}(S) - \sum \frac{|S_v|}{|S|}\mathrm{Entropy}(S_v) \tag{2.21}$$

where $\mathrm{Entropy}(S) = \sum_{i=1}^{c} -p_i \log_2(p_i)$, p_i the number of S instances belonging to the class i, A the considered attribute and $S_v = \{s \in S | A(s) = v\}$. Entropy measures the homogeneity (in which little entropy is sign of big homogeneity) of a group of instances, and as the tree decomposes into branches, it is of interest that the degree of entropy starts to decrease until a well defined class is reached. Figure 2.2 is an example.

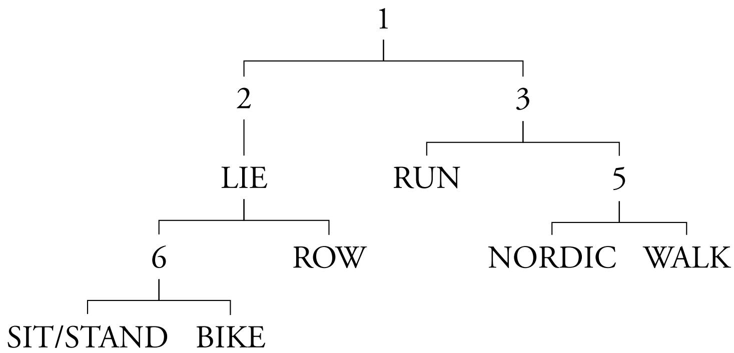


Figure 2.2: Example of a decision tree where an instance is evaluated, reaching a class in the final branch [34]

A main advantage of this kind of algorithm is its simplicity, meaning that once the model is created a group of if else blocks are capable of classifying every instance quickly and efficiently, however, its approach is not very robust which means that the model may be too adjusted to the training data used in its creation. Many ways have been tested to avoid this problem and one of them is to use a multitude of decision trees and consider the result to be, for example, the mode of the results. This principle is what supports the use of another commonly used algorithm, random forest.

2.4.2.4 Bayesian algorithms

These methods apply explicitly Bayes' theorem, a rule that describes the probability of an event based on prior knowledge of conditions that might be related to the event and it's mathematical form is:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad , \quad P(B) \neq 0 \tag{2.22}$$

with A and B as the events, $P(A)$ and $P(B)$ the probability of an event and $P(B|A)$ the probability of observing B given that A is true. If all the events possess the same probability the denominator is equal for all cases and therefore, only $P(B|A)P(A)$ needs to be maximized. Algorithms like naïve bayes, used in many cases for text categorization with word occurences as features, make use of this rule to classify instances and as advantage it is easy to implement with a linear time, but as a disadvantage assumes independency between variables which in some cases is incorrect.

2.4.2.5 Clustering algorithms

Unlike all other algorithms explained until now, clustering algorithms are a group of techniques of unsupervised learning, meaning that the training data provided to the technique has no label attached to it and the algorithm aims to group instances by their distance or similarity. Instances have big intra cluster similarity and little inter cluster similarity and even though the concept of clustering is common in all these techniques, the way each algorithm constitutes a cluster and how it finds them differs in each technique. Connectivity models like hierarchical clustering takes into account distance connectivity; centroid models such as k-means or k-medoids represent each cluster by a single mean vector; density models like density-based spatial clustering of applications with noise (DBSCAN) define clusters as connected dense regions in the data space, between others. Centroid models are between the most used techniques of clustering [33] whose k-means is an example. If k-means algorithm is to be summed up, it takes as input a variable k which is the desired number of clusters; randomly assigns k points on the data space and calculates the distance of every data point to each center using an euclidean distance metric by default. From the available distance metrics, euclidean and cityblock were tested where the latter is calculated by the formula:

$$d(p_1, p_2) = | y_2 - y_1 | + | x_2 - x_1 | \tag{2.23}$$

Which is very similar to the euclidean distance but where a border of points at a same distance make a square instead of a circle with a specific radius. After calculating the distances, each point is assigned to the cluster whose center is closer, with a partition made and followed by a recalculation of the centers (by averaging every x and y coordinate). If all data points remain on the same clusters, the algorithm stops, if not the process is repeated. Its easy implementation stands as an advantage, but as disadvantage is the lack of robustness to data with noise or outliers.

With the k-medoids algorithm centers are not attributed as random points in the data space but instead are considered as one data point, occuring the iterations by changing the chosen points for the centroids until the distance of every point to its corresponding center is minimized. This approach is theoretically more robust to noise, but with big datasets is not efficient. With the fuzzy c-means technique one point can belong simultaneously to more than one cluster, but is assigned to the one whose degree of membership is bigger. In subtractive clustering every point can be considered as a cluster center and based on the highest density of surrounding data points a first center is assigned and the neighbour points excluded; then the process repeats until all points are either centers or removed by a close center. Finally in hierarchical clustering, particularly in agglomerative nesting a distance matrix is calculated by getting the distance between every pair of points and the closest pairs are merged into a single point; the process keeps iterating until a pre defined k number of clusters is achieved. It is worth noting that as an unsupervised learning algorithm, clustering ends the process by grouping data without giving them a label or a true meaning, meaning that the user will have to, very likely, manually understand the relations and attribute classes posteriorly.

2.4.2.6 Dimensionality reduction algorithms

In some problems, a big amount of characteristics are withdrawn from the data, resulting in a inefficient and slow processing by the algorithms, moreover if systems with few resources and power are considered, a reduction in the number of the so called features is not only good from the point of view of processing but also from the problem's perspective since many of these features may contain redundant or not relevant information, worth to eliminate. Therefore techniques like principal component analysis (PCA) perform a linear mapping of the data to a lower dimensional space in such a way that the variance is maximized. In order to do this, the correlation matrix as well as the eigen vectors are calculated and the ones that correspond to the biggest eigen values are used to reconstruct the original data. However since the resulting features are a linear combination of the original ones, the final instances will no long have a physical meaning.

Chapter 3

State of the art

3.1 Commercial Trackers and Activity Monitors

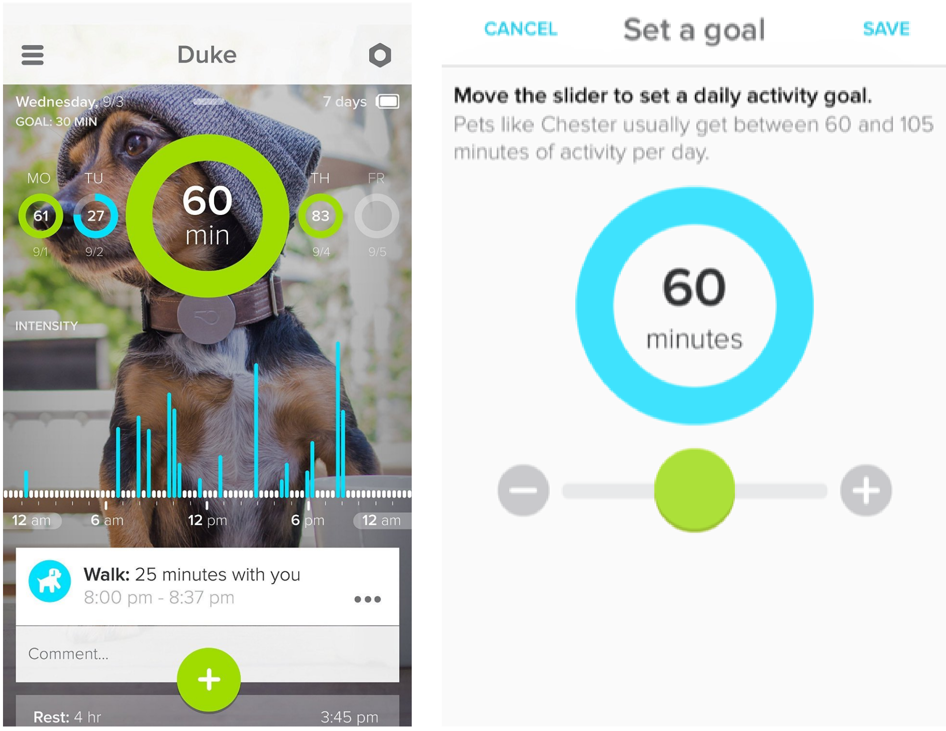
There are inumerous systems existing in the market that aim to solve the same problem as the one evaluated in this project. Therefore it is relevant to do a description of these available solutions and try to understand to what extent is this system better than the others, taking also into account that Findster's system stands as a different solution. Even though a complete list of all the similar solutions is not presented, since it is out of this project's scope, a representative group is shown considering and describing the most popular solutions.

3.1.1 Whistle

Whistle is a GPS pet tracker for dogs and cats, that uses a dog-collar to give location and smart monitoring in real time, in which the size of this device a downside. The techniques used to classify the type of activity and its correspondent level are machine learning or supervised learning techniques, very common in activity tracking systems for humans. In this system and like all the others, raw data is obtained and in the case of a 3-axis accelerometer three values are retrieved. The acceleration in the x, y and z axis for each instant. From the data, several features are obtained such as the mean or the standard deviation of the acceleration values in every instant, to use in an algorithm that tries to discover a pattern in each type of activity allowing it to predict which kind of exercise is the dog performing at a certain moment. The used algorithm is the support vector machine and it classifies, in dogs, 5 differente states, walk, run, play, active and rest. In other companion animals such as cats, the available states are only active and rest [41]. In this system, a timeline is presented, describing the dog's daily activity where a representation is made only of what is considered to be an event, *i.e.*, a block of at least 5 minutes of continuous activity. a reference is made to the fact that the sum of these events do not correspond entirely to the total activity performed by the dog once it can have fast episodes of exercise, lasting less than 5 minutes. After this activity event, the device needs to detect a rest period of at least 7 minutes so that another event can be counted.

A downside of this system, is the fact that the pairing of the dog collar with the phone is made using Bluetooth, a technology that can reliably transmit data with a maximum range up to 30 meters (however this transmission is dependent on many factors such as device orientation, enclosure, antenna design, etc)[42]. This means that an owner with a bigger distance separating him and his pet, will not have a real time feedback on the activity level, but instead will see the values saved upon the device coupled to the dog, and once the distance is reduced again, and the pairing restored, the data will be updated.

A special detail used in this sytem is that if the owner is in the bluetooth range for half (or more) of the dog's active time, Whistle registers a special note with that information on the activity event (a simple message saying that the animal was only active when the owner was close). Several owners are able to pair their phones with the device and obtain access to the records of the dog's activity (figure 3.1a). Once the pairing is done, it is also possible to infer with whom was the dog at a specific time [43]. There is also a "Daily activity goal" tab in format of text that takes into account the specific characteristics of a pet and presents a suggestion regarding the amount of active time per day, such as, "Pets like Chester usually get between 60 and 105 minutes of activity per day" (figure 3.1b). Along with this suggestion there is a customized goal of active time that can be set by the user. The information requested to the user is mainly the pet's weight, age, gender and breed [44].



(a) Main layout of Whistle's activity tracker **(b)** Activity suggestion and customized goals

Figure 3.1: Whistle's application screenshots [45]

Considering technical details, in this system the sampling frequency is 50 Hz, meaning that the accelerometer will provide three values of acceleration (considering a 3-axis accelerometer) every 1/50 seconds. Finally, and considering that techniques used in machine learning improve with growing amounts of data, an interesting feature is the fact that the user can contribute on improving the classification algorithms, since he can choose to correct a wrongly classified state by choosing the right one [46].

3.1.2 Fitbark

Fitbark is a solution more specialized in activity, leaving aside the GPS tracking part. However, it monitor's not only the pets' activity, but also their sleep quality where the activity information presented in Bark points, a proprietary unit. Other than these also a number of other interesting properties may be shown like the hours of sleep without interruption, post therapy recovery, etc. Its data platform is extended for use in academic and clinical research.

When it comes to the technology applied, patents protect their algorithms so there is no clear exposition in the literature, but considering statements like ”By continually grabbing data from accelerometers and feeding it to proprietary algorithms, these devices can generate surprisingly detailed information” or ”FitBark's algorithms are optimized for dogs, so using one on a cat or other animal—while possible—is not officially supported” [47, 48] may indicate the use of data mining techniques like classifiers that are improved with increasing data samples. Once again, in the smartphone you have the possibility to adjust a daily goal for activity (figure 3.2), with a goal not set in hours of activity, but in Bark points. Also like in Whistle, the dogs' characteristics are considered and three suggestions are given, where an average dog fits into the first, belonging in the 50th percentile of all dogs, an active dog into the second, in the 75th percentile, and an ”olympian” dog into the third, in the 90th percentile. There is also a menu that shows the the current number of daily, weekly and monthly Bark points [49].

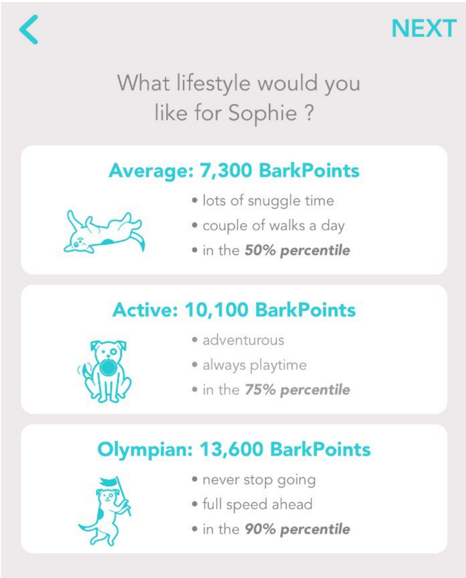


Figure 3.2: Activity goals suggestions in Fitbark [50]

3.1.3 Tractive

This GPS tracker also works with the same principles of those explicited before, with some interesting features. It also allows the user to define a circular virtual fence that is the safe zone, records a 24 hour history of the pets' location, it has a display screen embbeded in the device itself that shows the battery level as well as the number representative of the ”Petpoints”, another proprietary unit of activity that is updated in real time in the device accordingly to the animal exercise. However a great disadvantage in this system is that the GPS and the activity tracker are two different devices, making it hard to acquire information from both at the same time. The application in both gives the option to choose between one of four options of animal, dog, cat, horse or other and the coupling must be made with a specific collar that emits an alert if the device is removed (this feature only present in the latest version), a heat sensor that also triggers an alarm if the temperature around the dog is too hot, and an integrated light source to help find the dog in dark places [51].

Unfortunately no information was found about the technology used to infer the activity state or level.

3.1.4 Pod Tracker

This GPS activity tracking system has put a lot of effort on the instrument's design, where it is one of the lightest devices in the market. Their main advantages are the waterproof, impact resistant design, its interchangeable batteries and the possibility to enter on a "pet social network", with the creation of a profile for each dog, the possibility of comparing activity goals with others, etc. As a setback related to the GPS tracker is the lack of customization of safe zones, where the user can only establish circular zones for his or her pet. Concerning the activity tracker perspective, this alternative does not give many insights on the dog behavior since it only reports an estimation of walked distance, time spent in exercise and in rest, having no clear distinction between different activity levels.

3.1.5 Other systems

Systems like Kyon Pet Tracker are not in the market yet, but they have features obtained from different applied technologies.

This system is also composed of a GPS activity tracker but uses a LED (light emitting diode) screen to display text messages warning the owner or someone in a close proximity. According to the company, it incorporates a water sensor capable of warning the owner when the pet is at risk of drowning sending an alert, an altimeter that takes advantage of air pressure to infer in which altitude the dog is and more interestingly it incorporates a "dog fight pacifier" that emits a high frequency sound, when it detects an agitated state, forcing the dog to settle down for a while.

Finally to detect his activity as well as specific emotional states it uses a 9-axis inertial measurement unit that incorporates a gyroscope and a magnetometer, that once more, with proprietary classification algorithms can classify wether the dog is in specific states such as depressed or sick (figure 3.3). The biggest disadvantage is the product's high cost.

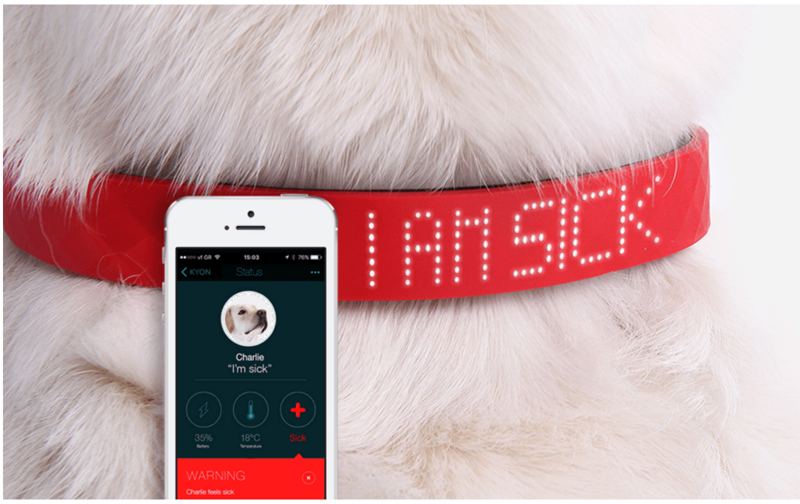


Figure 3.3: Message display in LED screen [52]

3.2 Experimental and Non-Commercial Solutions

In the research area an intensive search was made in order to understand what kind of algorithms are applied in systems that classify activity states and quantify practically the exercise made. As it was expected much more information was found when human systems were considered, which makes a division between human and animal systems suitable.

3.2.0.1 Systems Applied to Humans

An approach used many times is the attempt to create a linear regression equation that takes into account (therefore as an input) factors like, age, height or weight, together with the 3 acceleration values obtained from a 3 axis accelerometer and calculates directly the value of energy expenditure. In spite of its simplicity, this method is not robust enough once the equation has to be changed accordingly to the type of exercise made, lacking the existence of a single equation that can be applied for all activities Therefore an attempt to rectify this was made when a group of different equations was considered, putting the responsability of choosing the best suitable equation on the sensor, depending on the received raw data [53]. For this choosing to happen, features were extracted from the raw data in blocks of 4 seconds and once this block was classified as a specific activity, an equation from a set of five was chosen and the energy expenditure calculated. The workflow of this system may not be the best solution for the problem considered in this project, since the activity classification would have to be made in real time, and the processing power in the used device is low; even so, some relevant conclusions were obtained such as specific types of equations that present good results when used in certain human activities.

Actigraph's Freedson equation is the one that showed best generalized results, Hendelman equations were best to calculate EE in activities such as walking and jogging [14] and a possible manner of improving results is combining a set of two or more equations into one.

A tiny but interesting variation to the approach explained above is the use of not only linear equations but also models of exponential regression, used mainly in the calculus of EE in situations of more activity [35]. In the considered system, MET units are used, since they are better at the establishment of weight loss plans. In order to build MET regression models, key features were also taken from the accelerometer, more specifically from a parameter refered to as "signal vector magnitude"

$$\mathrm{SVM} = \sqrt{\mathrm{Acc}_x^2 + \mathrm{Acc}_y^2 + \mathrm{Acc}_z^2} \tag{3.1}$$

where mean absolute deviation, entropy or norm, are some examples of these features. Then these signal characteristics were fed into a software that through bivariate correlation analysis would select the most important features, to fit into the equations or models as an input [35].

Another approach and the one that is also more used in the commercial solutions is the use of classification or pattern recognition algorithms to infer which kind of activity is performed [38]. In these a similar workflow was found, it composed of:

- **Pre processing** - Mainly is when one or more filters are applied to withdraw the gravity acceleration or DC component of the signal

- **Segmentation** - Where a division of the pre processed data is made to a set of blocks with a specific duration that can be 4, 10 seconds or even 5 minutes, depending on the system.

- **Feature extraction** - The set of characteristics obtained from the signal that better ilustrate the different activity states.

- **Feature selection** - In most cases, the number of features is bigger than it should, a selection of the best ones is made or a combination of them takes place in order to reduce redundancy and incorporate all the valuable information in the feature set.

- **Classification** - Then these features are "fed" into a classifier algorithm that is capable of determine which kind of activity is taking place.

From the analyzed literature the most common features were the mean and standard deviation of acceleration in each of the three axis, the average acceleration, step count and time between two consecutive peaks.

3.2.0.2 Systems Applied to Animals

Even though there is not a vast group of projects with correct classification of animal physical activity, a lot of research is still made in the area. A note worth mentioning is that while commercial solutions focus in classification problems, with machine learning techniques used in the creation of a model that is "fed" with increasing amounts of data and, theoretically, shows increasingly better results (data oriented approach) experimental and non-commercial solutions focus statistical analysis and tests of the data in order to create a model that is robust enough to understand a relation between the obtained test data, wheter it is information from an accelerometer, from a camera, or other equipment, and the different activities practiced by the animal. This approach tends to be more ambitious once it tries to create a single, robust model with no need of capability to improve with more data.

All of the studies alert to the fact that the conditions taking place in a laboratory test vary substantially from the ones that animals have in their natural environment and if companion animals and kennel dogs are to be considered, their environment varies even more. Therefore an effort is made in order to mimetize these conditions, but still it is possible to understand that even if a solution presents good results in a predictable environment, there is still a huge step towards full validation of the method. Once again the gold standard technique to quantify or classify activity is indirect calorimetry, a logistically hard and costly method. If an equation or formula capable of containing information as parameters like age, gender, breed, weight, and calculating or expressing in a single value an activity made by an animal is to be found, at the time of this document writing, such equation does not exist. After some research it was possible to understand that what exists and with abundancy is the calculus of metabolic energy or rate, accessible by the owners to understand the number of calories needed by a companion animal so that it can survive, considering basic activities such as breathing maintaining body heat, digestion, etc [18, 54], which is equal to the already described resting energy requirements (RER). As an example, if a dog with a RER of 1175 kcal/day has a daily activity of 5 hours translated into a MER of 1878 kcal/day, that means the animal spent 703 kcal exclusively because of physical activity in that period.

Another way of quantifying the number of calories spent in a time period is making use of an index, indicating that every dog spends 1,1 kcal/kg/km if its velocity is approximate to 6 Km/h. This means that a dog with 25 Kg running in an hour will spend 165 kcal. If this energy spent in an hour is multiplied by 5 (to approximate to the previous example) this indicates the dog will spend a total of 825 kcal, a value not very distant from the obtained before. Considering also that losing 1 kg of fat corresponds to a energy spent of 7700 kcal, it would take 10 days for this dog to lose that amount of mass. A note worth mentioning is that both these approaches do not present a perfect precision, specially the second one since it is not coherently referenced, and so there are special factors to be considered such as the readjustment of the dietary regime which are not taken into account. When it comes to acquiring the data, two major instruments are used, accelerometer and pedometers. Some approaches make use of vests worn by the animal, composed by a group of accelerometers placed in different locations [55] and some use just a single accelerometer placed in a neck collar [56]. During a time period the values obtained by the acelerometer are accumulated giving origin to what is often refered as activity counts or activity counts per minute (if the time period is a minute). This measuring period varies accordingly to the projet that is taken place since it is not uniform. Some studies aim to take as much information from a test as possible and through that perspective, the measuring period is shorter (15 seconds most of the times), while other studies possess or gather better testing conditions, using a longer period per counting, such as 1 minute. Once the measuring is done, static levels, or as they are called in literature, cut-points are established to distinguish activity states such as sedentary, walking or troting [56]. A few examples, may be an activity count superior to 2000 counts/min typical of a vigorous activity state or an activity count inferior to 20 counts/min typical of a sedentary state.

3.2.1 Classifiers

A big set of classifiers were applied in problems as the one studied, so just a few of them are going to be depicted. Decision trees defined manually or automatically, in order to subdivide the problem into simpler questions is sometimes a good approach [34]. A practical example is when a state of running is taking place, the algorithm (initially) just assumes that the dog is in a vigorous activity rather than resting and only after, in a more specific branch of the tree, it tries to guess what kind of vigorous activity it is upon. This kind of approach can lead to less significant errors in the classification process.

Besides a note worth mentioning is the fact that algorithms based on decision trees can be applied in devices that rely on micro controllers, since just a few ”if else” blocks can group data in different activity types. In a first phase the conditions that separate the tree’s branches would have to be calculated on a computer, with a more significant processing power and once these conditions were obtained, the implementation on a microcontroller would be achievable.

Beyond decision trees, other techniques such as neural networks, support vector machines, k-nearest neighbor and others are also widely applied [34, 36, 40], but once again since the classifying process must be made in real time, and preferably on the modules, memory and processing power are limited and cannot withstand the complex calculations operated in those type of algorithms. A possible solution for the use of these models may be to incorporate the calculations elsewhere, maybe in a cloud or even in the smartphone once the module is paired with it.

Chapter 4

System Details

4.1 System Requirements

In this chapter the system's requirements should be described. Their division is made into six areas with one of four categorical levels, giving a perspective on the whole project's architecture.

The main goal is to develop an algorithm that can be implemented on a microcontroller and is capable of communicating three main variables through bluetooth to a smartphone application: the activity level, number of steps and spent calories.

The depicted requirement division is made with the following types:

- **Functional**: Describes system characteristics and what it should do;
- **Autonomy**: Describes capabilities that should be held by the system to ensure its autonomy;
- **Regulatory**: Describes the norms and directives the system should comply with and certificates that should be provided to the costumer;
- **Alarms and Errors**: Describes whose situations must trigger an alarm or error and what should be the system's behaviour;
- **Interfaces**: Describes interfaces with users and with other devices (type, range, protocol) ;
- **Non-Functional**: Describes the constraints on design or implementation, such as performance requirements, quality standards or design constraints, how easy is the software's use, how quickly it reacts and how well it behaves in unexpected conditions;

For each requirement there is a sequential number or ID for identification purposes and a list of affected requirements. In order to attribute an importance level for every requirement, one out of four possible categories is chosen, being these given by the MoSCoW method. This prioritization technique used in management and software development has an acronym derived from the first letter of each prioritization category, being those *Must have*, *Should have*, *Could have* and *Won't have*.

- **Must have**: Requirements classified as *Must have* are critical for the system's success so they must all be fulfilled during the project's given time;
- **Should have**: *Should have* label is for important but not necessary requirements. They can sometimes be as important as the previous ones, but not as time critical;
- **Could have**: *Could have* requirements are not strictly necessary while being desirable. They can improve user experience with little development cost and if time/resources are available, can be included;
- **Won't have**: *Won't have* requirements are the least-critical or not appropriate at the time. They may be dropped or reconsidered for inclusion in future work;

Table 4.1: Functional Requirements

ID	Type	Requirement	Category	Interfers with
F-01	Functional	The *pet module* must be capable of assigning a specific activity level to a dog's exercise in real time	Must have	
F-02	Functional	Correctly assign activity levels across different breeds (including mixed ones), ages, weights and gender	Must have	
F-03	Functional	Calculate number of steps of any given dog both outdoors and indoors	Must have	
F-04	Functional	Calculate dog's energy expenditure (in calories) in real time	Must have	
F-05	Functional	The *pet module* must be able to send calculated data through bluetooth notifications	Must have	
F-06	Functional	Each outgoing packet sent from the *guardian module* to the smartphone must not exceed 20 bytes (maximum length allowed by BLE)	Must have	
F-07	Functional	The mobile app must be able to manage data received from bluetooth notifications	Must have	
F-08	Functional	The mobile app should calculate dog's travelled distance in real time (both indoor and outdoor)	Should have	F-07
F-09	Functional	The mobile app should calculate the active and rest hours in real time (both indoor and outdoor)	Should have	F-07
F-10	Functional	The mobile app could save information about the user's dog such as weight, age and size	Could have	
F-11	Functional	The user could establish a daily activity goal and display its fulfillment in real-time	Could have	
F-12	Functional	The mobile app stores information on a daily, weekly and monthly basis	Won't have	

Table 4.2: Autonomy Requirements

ID	Type	Requirement	Category	Interfers with
A-01	Autonomy	Fast processing operations on the microcontroller to keep it on awake mode for the least possible time	Must have	
A-02	Autonomy	Least sampling rate that allows a proper signal processing	Must have	
A-03	Autonomy	The mobile app must receive and update data both when it is open and running in background	Must have	
A-04	Autonomy	Use of accelerometer's embedded high pass filter	Should have	

Table 4.3: Regulatory Requirements

ID	Type	Requirement	Category	Interfers with
R-01	Regulatory	The algorithm must be compatible with CE and FCC norms regarding RF communication	Must have	

Table 4.4: Alarms and Errors

ID	Type	Requirement	Category	Interfers with
A&E-01	Alarms and Errors	The mobile app must display a message if the bluetooth connection is lost	Must have	F-05, F-07
A&E-01	Alarms and Errors	The mobile app could notify the user when a daily activity goal is completed	Could have	F-07
A&E-02	Alarms and Errors	The mobile app notifies the user if his/her dog has not been active for a long period	Won't have	F-07

Table 4.5: Interfaces

ID	Type	Requirement	Category	Interfers with
I-01	Interfaces	The mobile app must be capable of clearly displaying received data such as the step number and spent calories	Must have	F-07
I-02	Interfaces	The mobile app should have a widget to present information on main screen, updated in real-time	Should have	F-07

Table 4.6: Non-Functional Requirements

ID	Type	Requirement	Category	Interfers with
NF-01	Non-Functional	Calculations on the microcontroller should be as simple as possible to reduce processing cost and time to a minimum	Must have	F-01, F-03, F-04
NF-02	Non-Functional	Preference should be given to variables that allocate little memory, for instance, choose 8 bit integers over floats	Must have	F-01, F-03, F-04, NF-01

4.2 System Architecture

Also an overview of the final system architecture is relevant with a compact description of every used component along with their communication protocols.

The accelerometer and microcontroller are incorporated into the same device, where the communication between them is through I^2C (inter-integrated circuit) bus, tipically used for intra-board communication with the attachment of lower-speed peripheral integrated circuits to processors or microcontrollers in short-distance. With it, a master, in this case the microcontroller can communicate with a slave, the accelerometer, to extract the acceleration in each individual axis. Once the accelerometer data is passed to the microcontroller, the created algorithm performs its calculations so that the number of steps, level of activity and spent calories are obtained and communicates via bluetooth low energy to the smartphone, as expressed in the block diagram of figure 4.1. During communication, the microcontroller is advertising a packet while the smartphone searches for nearby devices to connect, then, and only when the connection is established, the smartphone enables notifications on a specific service/characteristic allowing for data sending by the microcontroller every time values change (process better depicted further in 5.2.2)

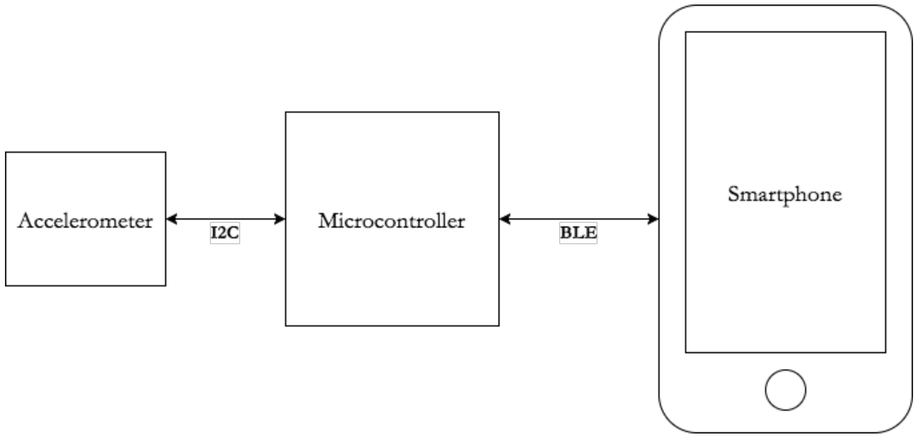


Figure 4.1: System block diagram

Chapter 5

Methodologies for Activity Monitoring

5.1 Conceptual Approach

First of all, there was a necessity of obtaining real, canine accelerometry data and for that a group of animals were tested with the provided systems in order to obtain raw accelerometer values on all sort of physical activities and postures by the pet.

During the measurements of this phase, the animal was holding a neck collar with the *pet module* (such as in figure 5.1), for real time step quantification in a time window with a specified length.

Figure 5.1: *Yorkshire* wearing a neck collar with the *pet module*

5.2 Technologies

5.2.1 Three-Axis Accelerometer

The accelerometer in use for this project is the NXP's MMA8652FC 3-Axis, 12-bit, micromachined digital accelerometer. As its main useful features, it possesses embedded low pass and high filters that can assist on signal processing, more specifically in distinguishing the static acceleration caused by gravity and the dynamic acceleration originated by the animal's movement. It has different scale ranges, where the one in use 2*g*, which corresponds to a sensitivity of 1024 counts/*g* and it contains an internal 32 sample data buffer, which helps minimize the number of time the accelerometer has to communicate with the microcontroller since processing can occur after 32 samples are collected instead of processing every sample each time. This internal buffer or FIFO can be configured to work on one of three methods, collecting all the 32 samples and notifying when the buffer is filled (fill buffer mode), allowing for constant data entrance and when the buffer is filled, new values start to replace the older ones (circular buffer mode) or can keep data when explicitly told to (trigger mode). For sampling frequency 4 options are provided from 1.56 to 50 Hz, where the one used in this project 12.5Hz. This choice was made to create similar conditions to the ones in the already created system used by Findster and tries to encompass a sampling rate capable of describing the animal movement while having low power and processing costs. With regards to the general functioning of the accelerometer a few guidelines must be taken into account, namely:

- Returned data from the accelerometer must be converted to *g* units, more specifically *mg*, different from the gravitational constant G. Even though it is not used by the international system of units (since it uses g from grams) it is the most widely used in acceleration related problems.
- To convert the returned values to *g* units they must be multiplied by $\frac{2}{2048}$ or $\frac{1}{1024}$ since a 12 bit can represent $2^{12} = 4096$ different values where 2048 and -2048 are representative of 2*g* and -2*g* respectively.
- One *g* is defined as the acceleration due to gravity at the earth's surface and is equivalent to 9.807 m/s^2.
- Other than the constantly applied force of gravity, the device is subjected to external movement caused by the oscillation of the animal's body and the acceleration that comes from it, is due to a vector sum of all non-gravitational and non-eletromagnetic forces acting on the device.
- When the device is stopped, 1*g* of acceleration is applied in the upward direction, because it is keeping the object from going into free fall, ensuring that it is accelerating relatively to this free-fall condition into the earth's center. Therefore when one axis of the accelerometer is alligned with gravity force, it is theoretically expected that a value close to 1000 (*mg*) is returned, with this signal component also called DC since it sets the average value of the signal different from 0.
- If the accelerometer is tilted relatively to the previous position, it is expected to see a distribution of the 1000 *mg* g-force for two or three axis, where the absolute value (obtained as in equation 2.8) on all three axis are similar.
- Dynamic acceleration caused by the pet oscillation, or AC component is the unpredictable part of the signal to be considered. When raw data is considered, this component is attached to the static acceleration, creating a necessity of separating both these

elements and one way to do it is recurring to the frequency domain where the embedded digital filters can be of use.

5.2.2 Bluetooth Low Energy - BLE

Bluetooth low energy or bluetooth smart is a wireless technology that gives mobile applications access to external hardware. It differs from classic bluetooth since it does not maintain a constant connection but instead only sends data when needed which is a good option for use cases such as sensor readings. BLE has only one packet format and two types of packets, advertising and data. The former broadcast without the knowledge of a scanning device and sent at a fixed rate. A BLE device can communicate in one of two ways, broadcast or connection and in the latter a permanent, priodical and private data exchange of packets between two devices. To establish a connection, a master scans for advertisers that are accepting connection requests and when a slave is detected, it sends the request packet and after that connection interval the number of events a slave can skip without disconnect and maximum time between two data packets is sent by the master and acknowledged by the slave. These connections involve two separate roles, a central or master that scans preset frequencies for connectable advertising packets and initiates a connection and a peripheral or slave that sends advertising packets periodically and accepts connections. Once the connection is established, the peripheral stops advertising giving space for the data exchange between the two devices in both directions. Connections have more control over data since it is organized in services and characteristics (services can contain multiple characteristics) by an additional protocol layer, the Generic Attribute Profile or GATT (used in the created mobile app). The attribute protocol is a client/server protocol where the client requests data from a server which must send it before any other request can be made. Each server contains data organized in the form of attributes that have a 16-bit handle, a universally unique identifier (UUID) that specifies the data's type, permissions and a value. If a client is to read values from a server it makes a read request to the server's specific handle, receiving an acknowledge letting the client parse the data based on the UUID [42].

5.3 Data Acquisition

5.3.1 Pedometer

In order to build a system capable of quantifying a step number in real time, as stated earlier, real canine data was needed and therefore a measuring method was applied.

The main components of this first measuring system were the *pet module* also referred as the TX module (given that it was the transmitting module), the *guardian* which served as a gateway between the *pet module* and the smartphone, receiving the accelerometer data by RF and sending it by BLE and finally a smartphone application responsible for acquiring data from the *guardian* and displaying important information such as the number of steps, activity level, the type of digital filter applied or even the values of activity state thresholds.

During a measurement, the TX module was placed in a dog's neck collar with button to start the data emission. Then and once again, only with a simple button press, the *guardian* (held by the person responsible for the measurement) started to receive this data and blinking the LED inside of it, where this waa sign that the system was, indeed, working. Lastly, on the smartphone application, the user was presented with a search menu that looked for bluetooth devices, like in figure 5.2a and presented them in a list where each row could be clicked. After selecting the TX module, coherently labelled ”findster”, the bluetooth pairing was made and a final screen was introduced where the user could start, pause and stop a measurement.

5.4 Smartphone Application

Unlike other components of the measurement system, the smartphone application was entirely built during this work, with the initial intent to only operate as a data displaying tool, but later used also as a facilitator of the measuring system. It was built using Android Studio, the official integrated development environment (IDE) and a SDK or software development kit tools that allows the creation of applications for a certain software package. This necessity of projecting calculations on the mobile app arrived from the fact that both memory and processing resources are scarce in other system components given that they rely on microcontrollers. With this architecture more complex calculations can be made and more values can be stored confering more liberty to the system, essential in this project's phase since reliable data must be acquired.

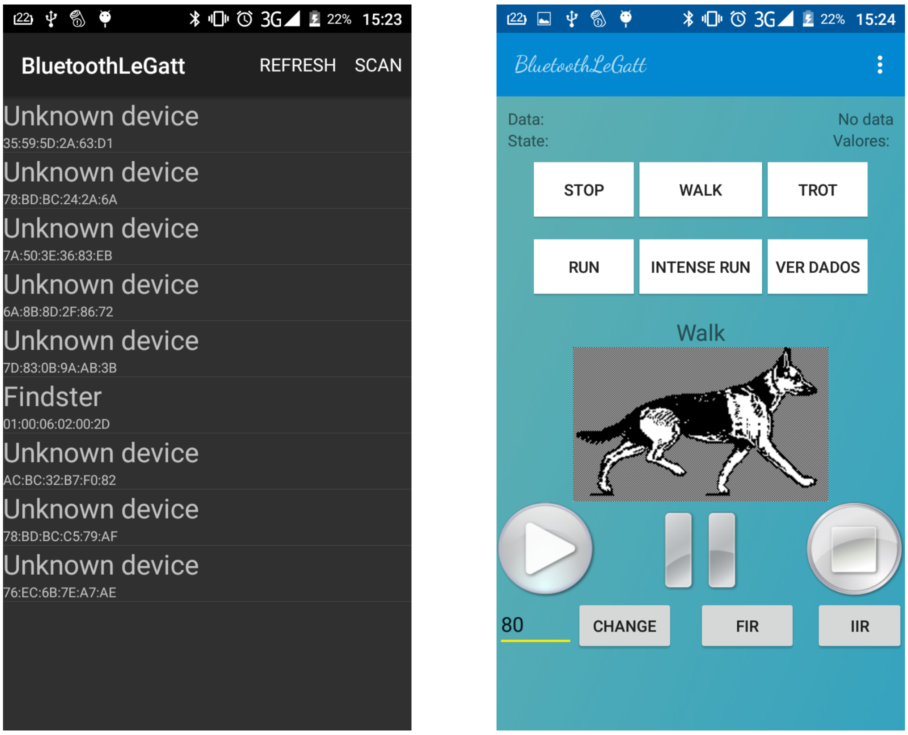
5.4.1 Data Logging

One of the main purposes of the application, was to receive accelerometer data and store it into a file that could be accessed later in other platforms for easy processing. The used solution should write the received values on a plain text file, with ”.txt” format and save it in internal phone memory. To do so a few challenges were encountered and different approaches were tested in search of an efficient algorithm that could be able to write data logs of long acquisition times without harming the efficiency and speed of the application.

In a first approach, the app wrote on a file but to do so, it loaded the whole content of the text file to a variable, adding only then the received values. This means that a big stream of data was loaded at a very fast rate. It is worth mentioning that every 80 milliseconds (12.5 Hz of sampling rate) at least a group of 3 values were delivered to the smartphone, and alongside them, a time-stamp was added to keep record if the values were arriving accordingly to the sampling rate. Ideally the time stamp was supposed to be communicated by the TX module to the *guardian* and this time stamp was precisely when the value was obtained, but adding those extra bits of information to the communication could cause a small delay on the transmission compromising the data's validity. Therefore the registered time-stamp is descendant from the application, corresponding to the time when the sample values reached the smartphone. The written time stamp stands as a very long number representative of the number of miliseconds between current time and the first of January of 1970, that is written on the same line as the acceleration values, separated by a comma. Returning to the logging approach, it was possible to infer that this method was not plausible since in one second the phone had to write, approximately 12 text entries (12.5 samples per second) and after some minutes the size of the text to be loaded into a variable was too big for the phone to handle, giving rise to an increasing lack of response, culminating into a complete freeze that originated a ANR or Application Not Responding. In an attempt to fix the described, a method that only searched for the final position of the file was applied, writing in that position the received data on a background thread so that the logging process would not affect the performance of the main user thread, where displays were made. This was accomplished with the use of an object called ”AsyncTask”, a class that allows the performance of background operations asynchronously without having to explicitly manipulate threads or handlers. Even so, this approach was not satisfying since the problem was mainly on the elevated number of readings of the text file, which often lead to, once again, the decreasing performance of the app over time. Finally, reducing the number of times the app had to access the text file for saving the data, by only doing so once every minute, showed improved results and turned the application adequate for long time measurements. In this approach the data is stored in a local buffer (a "String" variable) during 60 seconds and only then written in the document and giving space for the variable's content to be erased. In the case of a stopped measurement between writings, the "app" automatically is forced to store the variable's content in the text file, ensuring that no information is lost. Also if the smartphone enters in sleep mode the application is capable of keeping the record going thanks to a core android component called broadcast receiver that enables applications to receive intents that are broadcast by the system or by other applications even when the app is not running.

The final menu, such as the one in figure 5.2b, had five different buttons with clear labels referring to the 5 activity states that the system is trying to differentiate (stop, walk, trot, run and intense run) and their purpose was for the user to correctly identify what the animal was doing by pressing the correspondent button. This would not only compel the user to be envolved in the process, but particularly help the posterior signal analysis, because conclusions on how the signal differed in time windows of intense or no activity could be backed or refuted by those labels. This labelling process also represents a surplus if a machine learning is attempted in the differentiation of activity segments, since for the training phase, correctly labelled data is needed, however it is important to note that this *a priori* activity classification serves only as an auxiliary tool and should not be used as a ground truth, due to the fact that activity classification is in many times subjective and prone to user error. A few examples can be shown to prove that in many scenarios the user classifies wrongly the animal's movement such as when he or she loses sight of the animal, where is hard to assert which of the buttons to press; when the animal is walking or trotting, the frontier that separates these two states is by far the most hard to assess, meaning that for some users the same dog with the same movement may be considered to be walking by one user and trotting by another; also some users may simply forget to press the buttons leading to a inconsistently big time window in a level, or may take a long time to identify the level change, reflecting it on a delay in the labelling, etc. First the approach on writing these tags was to create a new line on the text file each time the user pressed a button with a simple word surrounded by "%" symbols so that the parser could interpret them, however this approach showed to be complex while implementing the parser and consequently a new approach was implemented with the writing of a discrete value on the interval from 0 to 5 in the same line as the axis aceleration values, resulting each time a value was received in the writing of 5 values, the accelerometer values in x, y, z, the time stamp and the correspondent tag for the activity level. Values 1 to 5 correlate to the 5 defined activity states (from stop to intense run) and the value 0 corresponds to an unlabelled state, normally wrote before the user pressed any button.

The application was conceived and designed with the intent to be as intuitive and simple as possible, so that it could be used for any individual owner of a dog. Accordingly, a user's manual (see appendix A) was written to help dog owners make use of the system and record several activity measurements of their companion animals, where log files were later provided for further processing.



(a) Search menu with bluetooh devices **(b)** Final menu with state buttons

Figure 5.2: Screenshots of the developed application

5.4.2 Participants

For the first phase of measurements, done only with the accelerometer, the population consisted on a total of 25 dogs with different ages, weights, breeds and genders. All of them presented no illnesses or difficulties in executing activity of any type, from slowly walking to extreme running. The owners were asked to record their dog's activity while clearly identifying the activity type as one of the five pre-defined types. In an ideal case the owner was asked to record different files for each of the activity levels, and in some cases that was achieved for the stop and walk activity state, but for trot, run and intense run, it was impractical to keep the dog at that level for long periods. During measurements made by the author, video recordings took place, with a digital camera, Nikon Coolpix L29, for further step accounting, so that numeric accuracy of the algorithm could be established by comparing the real step number and the algorithm's result. The ranges in weight and size were broad, ranging from 2 to 35 kg, with ages from 6 months to 14 years and different breeds, with the majority as mixed ones.

5.5 Methods

Once data was obtained, it was analysed with Matlab, an academic software used by engineers and scientists in a wide range of areas such as health monitoring devices, safety systems, etc. Previous working experience and possible applications in machine learning or signal processing represented relevant advantages for the choice of this software to the project.

5.5.1 Pre Processing

The majority of accelerometer (absolute) values are located in the range of 0 to 1500 mg, and once again the choice of $2g$ for the acceleration limits is due to the margin provided by the device. After ensuring that the signal is in good conditions, it was possible to infer that in some periods a baseline is present, happening when the disposition of the device is changed in the dog's collar. This position change of the device is an event that can be minimized with the tightening of the collar to fix the *pet module*, but inevitably with strong movements the animal is capable of moving it, with consequences similar to the ones presented on figure below.

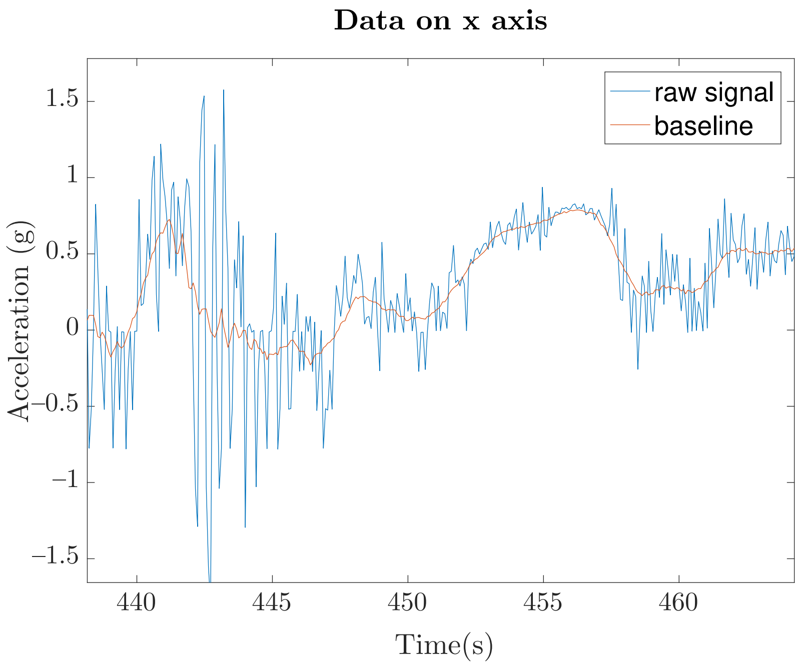


Figure 5.3: Raw data with baseline (orange)

To avoid the artifact present in signals like the one in figure 5.3, filtering the signal, namely its low frequency components, may seem a logical approach and for that a signal analysis in the frequency domain is needed, achieved easily by a Fourier's transform. Joseph Fourier, a french physicist and mathematician, in 1822, demonstrated that any signal could be decomposed in a sum of harmonics or synusoidal functions, giving rise to the formula.

$$x_n = \frac{1}{N}\sum_{k=0}^{N-1} X_k . e^{i\,2\pi k\,n/N} \tag{5.1}$$

where x_n is the value of the signal at time n, k the considered frequency (0 Hz up to N-1 Hz), N the number of time samples and X_k the amount of frequency k in the signal. Therefore when the Matlab function "fft" or Fast Fourier Transform is applied to the data, it is possible to obtain a spectrum of the signal's frequencies, represented by the x axis and their occurence on y axis (figure 5.6). In order to remove low frequency components as well as the DC component two types of filters were tested with a cut-off a frequency of 0.1Hz (also used in other projects [57]), chosen with the intent of not deleting important information.

5.5.1.1 Finite Impulse Response filter

FIR or finite impulse response filter as the name indicates is a filter whose response to an impulse or any finite length input is of finite duration. If a Nth order discrete time FIR is considered, its response lasts for N+1 samples before achieving zero and the output of this filter can be calculated by multiplying a set of N+1 coefficientes and the signal with unit delays, as expressed in the block diagram of figure 5.4

$$
\begin{aligned}
y[n] &= b_0\, x[n] + b_1\, x[n-1] + \ldots + b_N\, x[n-N] \\
&= \sum_{i=0}^{N} b_i\, x[n-i]
\end{aligned}
$$

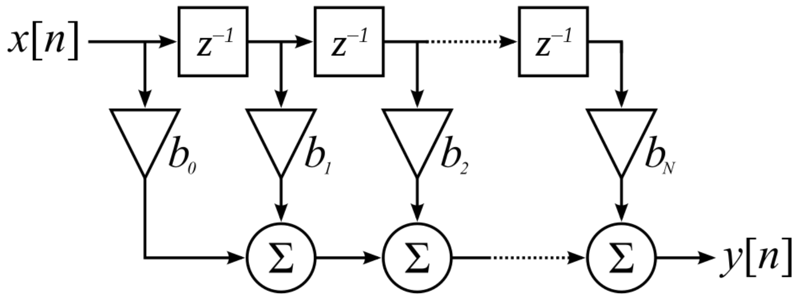


Figure 5.4: FIR block diagram

where x[n] is the input signal, b_i the filter coefficient or value at the ith instant, y[n] the signal's output, N the filter's order, and the calculation very similar to discrete convolution. As their main advantages these filters can easily be designed to have a linear phase response, meaning that the delays to the input signal will not distort it; they are easily implemented, done in most cases by looping a single instruction; since no feedback is required, some calculations may be ommited when for example a reducing in sampling rate is required, this a computational efficiency asset and they can be implemented using fractional arithmetic, *i.e.*, all filter coefficients may (and should in most cases) be less than 1. As their main disadvantages, frequency response is not as easily defined as in IIR (infinite impulse response) filters; in some cases requires more memory to achieve a given response obtained with a smaller order IIR filter.

5.5.1.2 Infinite Impulse Response filter

IIR or infinite impulse response in contrast to the previous one, has an impulse response which does not become zero after a known time, but continues indefinitely. This type of filter possesses memory and feedback that grabs previous outputs to define the current signal output as in the block diagram of figure 5.5

$$
\begin{aligned}
y[n] &= b_0\, x[n] + \ldots + b_P\, x[n-P] - a_1\, y[n-1] - \ldots - a_Q\, y[n-Q] \\
&= \sum_{i=0}^{P} b_i\, x[n-i] - \sum_{i=0}^{Q} a_i\, x[n-i]
\end{aligned}
$$

where P is the feedforward filter order, Q the feedback filter order, b_i the feedforward filter coefficients, a_i the feedback filter coefficients, x[n] the input signal and y[n] the ouptut signal. As its main advantage, is the efficiency upon implementation, meaning that the same response obtained from a FIR and IIR filter, generally is obtained by a lower order Q than in a FIR (N), which translates into fewer calculations per time step. If some phase distortion is tolerable or unimportant an IIR filter is preferable [58] given that less memory and lower computational

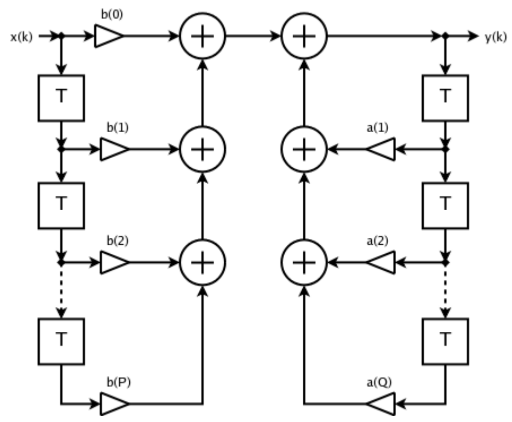


Figure 5.5: IIR block diagram

complexity is needed. As its main disadvantage is the complicated implementation once several loops and buffers are needed to encompass not only the unit delays of the entrance but the previous outputs. In order to eliminate low frequency components, both these types were tested with different number of coefficients and the main reason for this was to understand if in the case at hands, possessing no limitations on resources produces much better results than a system with scarce resources both in memory and in processing, a reality present both if the algorithm is to be implemented on a mobile application or in a microcontroller (with bigger limiations in the latter).

Two high pass FIR filters were tested, 5th order and 86th order (figure 5.7a). The choice for 86 as the order of the second filter was based on a Matlab's toolbox ("fdatool") option, "minimum order" that calculates the minimum order to have a filter cut-off of -3 dB at 0.1 Hz. Replacing the 5th order FIR filter with an IIR filter of 7th order (figure 5.7b) was applied for further comparisons, and the reason for not existing a test with a bigger order IIR filter is related to what was said earlier, that IIR filter can achieve good responses in the frequency domain with lower number of coefficients. With the "fdatool" it was possible to choose between some different FIR filter types such as equiripple, least-squares and IIR filters like "butterworth" or "chebyshev". The choosing criteria was based on the frequency response, showed also by this toolbox when the filter that could reach the magnitude of -3 dB with a frequency closer to 0.1Hz was chosen.

Finally the time vector was obtained by subtracting all values with the first time stamp in order to have a representation in seconds.

5.5.2 Processing

After ensuring that the data values provided by the smartphone application were in proper conditions, the main steps of analysis took place. First a quick test was made to the time vector, with the subtraction of every value by its predecessor, and this operation was similar to a first derivative and with the goal of checking if all values were spaced equally by 80 miliseconds, the time correspondent to the sampling frequency. Then a mean was applied to the difference vector and in the majority of cases the mean time per sample was in the range of 80-100 ms with a jitter of approximately 30 ms. Jitter is the deviation from true periodicity, which in the case at hand does not represent a critical issue once a dog cannot perform a big amount of movement during that time.

Following the verified tendency of many works done with accelerometry, the acceleration

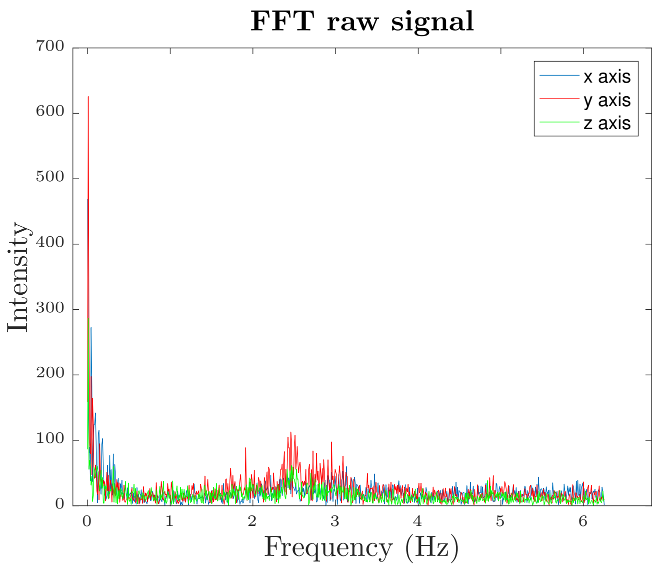
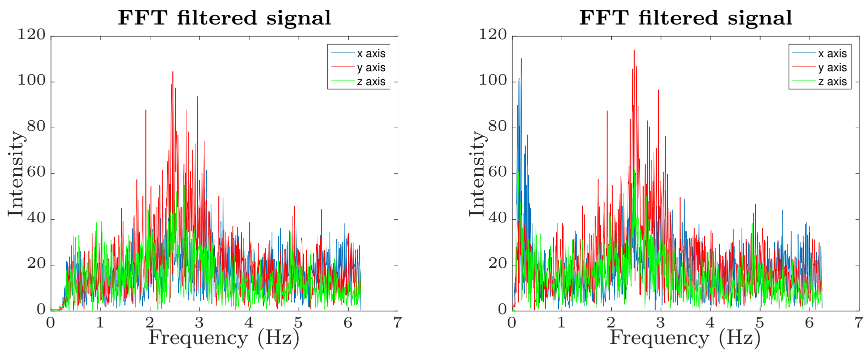


Figure 5.6: Signal's fft on 3 axis



(a) FIR filter of 86th order

(b) IIR filter of 7th order

Figure 5.7: Different filter types applied with a cut-off frequency of 0.1Hz

vector, (signal magnitude vector or movement intensity) was calculated for further comparison to the individual values of each axis. The first challenge was to analyse the signal in order to identify and match a dog step. For it, short measures and manual correspondence were made attempting to clearly describe a typical dog step morphology in the accelerometer signal. Theoretically it is comprehensible that when a quadruped animal performs locomotion, in any of the previously defined activity types, there are two main phases, an ascending followed by a descending one. In the first phase, the animal projects himself upwards obtaining a velocity increase that reflects in a dynamic acceleration, recorded by the sensor, until a maximum height point is achieved and both the velocity and acceleration become zero, then the descendant phase takes place and the velocity increases in the opposite direction while the height decreases, with an acceleration recording until the final moment where the animal reaches the ground. However, empirically, it was possible to understand that the acceleration recorded in different activity states from a slow pace walk to an exhausting intense run is mainly marked by the step's last phase, where the paws touch the ground and the whole body absorbs the fall impact, which is responsible for the biggest mass displacement on the accelerometer. A typical step is then translated into a signal composed of two peaks, a maximum and a minimum and in order to understand the main reason behind the influence of the fall impact, a short explanation on the accelerometer's functioning principles is relevant.

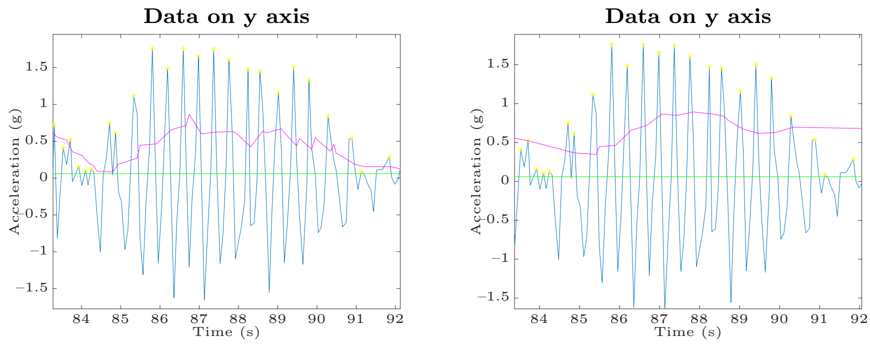
5.5.2.1 Accelerometer mechanism

Accelerometer sensors are based on the calculation of acceleration by the use of Newton's second law of motion (F = m a), which means that in fact, what they measure is a force applied to a known mass and only indirectly obtain the acceleration values by the displacement of a mass that is placed along with a mechanical spring between two fixed electrodes [59]. This displacement is caused by acceleration and it causes a change in capacitance that is detected and does not change with factors such as temperature, making it a reliable and highly accurate sensing mechanism.

Taking into account that the device's accelerometer placed on the neck collar measures mainly a velocity variation in the dog's neck and head, it is likely that when the animal reaches the ground, the impact is what causes a bigger mass displacement. Following this interpretation, an approach used in some projects (specially actigraphy) of using the absolute values was discarded, since each step and its descendent phase with negative values would be turned positive making each step characterized by two maximum peaks instead of one. An algorithm that could identify all the peaks in the signal was needed with the main purpose of understanding if their number of occurences corresponded to the overall number of steps. In Matlab this was accomplished easily, using the "findpeaks" function, but it was possible to infer that the number of peaks was always superior to the number of steps due to the fact that even when the dog is completely stopped, acceleration values suffer little oscilations originating peaks that should not be considered. This phenomena creates a necessity of conceiving a threshold capable of only counting peaks correspondent to animal steps and leaving aside noise related ones. In a first approach, several values were applied attempting to build a static threshold, but this method showed to be unreliable when faced to different dogs, when size and age factors influenced the way body oscillates. Another identified issue with the static threshold was the step counting in both intense activity levels and quick changes between states. In those cases the signal becomes unstable and several peaks per step are originated instead of only one and with that in mind, a second approach was to build a dynamic and adaptive threshold that would in a way adapt to the conformation of the data so that quick changes were smoothened and the number of steps would not overshoot. To build this threshold two major things were taken into consideration, a static baseline from which below, no steps could ever be counted and the second a weight vector that was multiplied by a variable number N of previous peaks. The threshold in each instant was obtained by the influence of these N previous peaks creating in a way a moving average to establish the threshold. The first implementation of this algorithm was made in Matlab and the were:

- considering only the peaks whose value is superior to the static baseline
- calculating an average sized peak by multiplying two user defined inputs, the number of previous peaks and its correspondent weight vector. These weights give different importances to peaks closer or more timely distant
- establishing the threshold value for that instant as half of the height between the previously calculated peak and the static baseline
- counting the peak as a step if its value is above the calculated value or discarding it otherwise.

An interesting variation on the described algorithm is held on the choice of what peaks should be considered to calculate the threshold value in a given instant. Two possibilities were tested where in the first every peak that stood above the baseline (above or under the threshold) was considered for the threshold's calculation, resulting in a more oscillating threshold (figure 5.8a), and on the second only if the currently analysed peak was bigger than the threshold in the previous instance, could it be used in threshold's calculus (figure 5.8b). Both approaches represent significant differences since in one, the threshold is much more data dependent, presenting oscillation in more fluctuating signals opposing to the second approach were a smoother line is obtained as it can be seen in the figure below.



(a) First approach on threshold calculation

(b) Second approach on threshold calculation

Figure 5.8: Same signal excerpt with the threshold accounting more (left) and less peaks (right), green and red line stand as the basal and dynamic threshold

In addition to the clear smoothening effect of the second approach, in both cases it is possible to identify a small delay of the threshold when compared to the signal, an expected phenomenon since each threshold value is calculated based on the peaks of previous instants, an effect also occuring when moving average filters are applied to any signal. The algorithm can also be tuned with the number of previous peaks considered (equivalent to the sample number in moving average) for the calculation of each threshold value and the weight vector associated with it. Several tests were made with the intent to discover the influence of a threshold with more memory, *i.e.*, that considers more previous peaks (from 3 to 50), and also the influence of a weight vector that puts more emphasis on closer or farther peaks. Finally, after applying the created thresholds to the signal of every independent axis as well as the overall acceleration vector, a set of metrics were needed to merge the step counts of the 3 axis, unlike the vector where steps can be interpreted directly as the peaks above the threshold. Considering that the count of only one axis is not valid because the device is not always placed in the same position, it is impossible to guess which axis should be considered (after signal's high filtering) and moreover, the device can change position during the measurement, meaning that the sum of all counts clearly overestimates the number of steps when the device is for example, on a tilted position. Using a mean is also a poor choice considering that in most cases, step counts are focused on one axis only, having the other two a very smaller value, pushing the final step value to an underrated approximation. Another alternative is to go with the maximum step number between the three axis, which in some cases showed to be a good alternative, considering that in short time measurements even if the device's orientation changes, has negligible effects on the final counting. Although, if during a long time period the device is alligned with one axis, lets us consider the y axis, the correspondent step number starts to increase, but once the device rotates and alligns itself with the z axis, the step number related to this axis has to reach the step number correspondent to the y axis for the final step number to change. This latency in response, makes the approach not valid. Therefore and considering what was said, the oscillation of dog's body happens in, at least, one of the 3 axis, meaning that when a peak with height superior to the threshold is detected in one axis, in two or in even in all axis, it should be accounted in the same way. This kind of metric is implementable with an operation similar to the bitwise OR, when a step is equivalent to a one (1) and a non-step to a zero (0). In each instant, one axis possessing a step (1) is enough for it to reflect on the final step quantification, making this approach robust to device's orientation.

With all previous methods, limiations on peaks' height were established but during measurements it was possible to check that a step overcount was taking place, specially when the device changed orientation on the collar, meaning that, for example, 10 steps were considered in half second which is physically impossible. This creates a necessity to limitate the signal peaks not only in height, but also in terms of time (x-axis in figure 5.8) in a way that typical dog gait limitations are respected. An empirically minimum time for a step of 100 milliseconds was established, meaning that two steps must be separated by this time feature which was set as an input for the user to change. Even though this time limitation turns the algorithm more robust to movement artifacts like device orientation change or spontaneous scratches of the animal towards the device, in those situations still some steps are accounted given that the accelerometer is not capable of differentiating walking movement from other artifacts. Moreover one big limitation of this approach is the fact that 100 ms is a delay representative of a dog moving with a big velocity, meaning that in slow pace movements the algorithm still is prone to detect anomalous peaks, if a bigger delay is introduced, the fast pace movements will display less steps than they should meaning that for the establishment of a permanent step delay, this balance has to be taken into account.

5.5.3 Activity Qualification

The work's main goals stand as the quantification as well as qualification of physical activity, *i.e*, the attribution of one in five pre-defined values for every time window. To do so, several parameters can be obtained from the now processed signal (explained in 2.4.1), but in a first approach, mean acceleration (2.4.1.1), often referred to as activity count is calculated for a time window. To establish thresholds that were capable of separating 5 different activity states, it was taken into account the ones used in previous works, where values like 562 and 2912 activity counts per minute separated rest and sedentary behavior from vigorous one [60] or 42, 287 and 776 counts per 15 seconds were used to identify stop, walking and trotting behavior respectively [56]. However these threshold values showed, through testing, not to be effective in classification and therefore, empirically, new ones were established. As an easy way to tune these values, the smartphone application displays an image related to the activity state (figure 5.2b), giving real time feedback and allowing the user to understand if the states were in fact well classified or not. During first trials, measurements were made with established limits for different activity states and through the result, the author had to change values accordingly, but as an upgrade, the mobile application was provided with a settings menu where the user is able to change limit values in real time, having a faster and more efficient tuning system. It was implemented with the use of an Android interface, "SharedPreferences" that not only changes parameter values but saves them between application shutdowns. Nonetheless, it may be of good use to note that activity counts cannot be considered equivalent to velocity because instead of integrating acceleration over a time epoch, a sum is performed on individual samples and moreover each sample is filtered before any analysis. This filtering process differs in every equipment and that can be one of the reasons why cut-points established in previous works did not match the ones in this project. Even so, previous works establish thresholds with the two most commonly used monitors in physical activity research, the Actical (Philips-Respironics, USA) and ActiGraph GT3X (ActiGraph, USA).

A not yet discussed matter is the choice of the time window's length. Its selection may not influence activity levels once the considered value is the mean acceleration and in the first implementation, this time length was equivalent to 10 seconds, result of the communication between Findster modules happening with that interval, however, if an adequate tuning is needed, the system needs to return a feedback faster enough to let the user correspond the displayed image with the current animal activity. Taking this and the sample frequency into account, a time window length of 2 seconds, correspondent to 25 samples was implemented. Also a note worth mentioning is that with this method not manual user labels nor complex signal features are used, privileging the simplicity and practical calculations for the activity qualification.

After some tests, and a tuning process, the system was able to properly classify activity levels such as walk, trot or run, but this calculation was made completely independent of the previous step counting. Even though that was not the case, the referred can lead to a data inconsistency like, for example, identifying an activity state of a stopped dog, but an increasing number of steps, which is not coherent. With the intent of solving this issue and also improving the step counting algorithm a simple relation can be made, where the activity state sets the minimum time interval between steps. This means that if the mean acceleration corresponds to a stopped state, no steps can be recorded, if the dog is walking a minimum interval of 200 milliseconds is implemented, if the dog is trotting, this interval decreases, etc. The implementation of this adaptive value and its testing led to important conclusions such as the time between steps, given that in a first approach it was decreased alongside the increase in activity state, but with testing that would result in an adequate performance while the animal was stopped, walking or trotting, but in a poor performance when peaks of activity, typical of a running dog took place, with the system returning anomalous step values. This phenomenon is strictly related to the dog's gait (see 2.1.2.1) and it's body oscillation in different activity states. When walk and trot are considered, the dog oscillates its body with every step and the steps per unit of time are increasingly bigger, but when the animal starts to run, it supports all of its weight alternatively on the forelimbs and hindlimbs, meaning that per stride, the body oscillates half of the times, but with higher and longer intensity. Therefore when a dog is running, the accelerometer is only capable of recording one step per two real ones, since they happen almost at the same time. As the animal's velocity decreases, the interval between steps increases until the acelerometer is capable of detecting, once again, each step. In practical terms this translates into a decrease of step interval until the trotting state, but once the dog starts to run, the step time should increase given that the time between the forelimbs and hindlimbs support is bigger.

Summarizing, in order to obtain a step number, with a step as the instant when one of the forelimbs touch the ground different methods were tested:

1. counting the number of times the signal crosses the value $0g$,
2. counting the maximum number of steps within the ones given by the independent axis
3. counting the steps with the "OR" method, and considering all peaks above the baseline for calculation
4. counting the steps with the "OR" method, and considering only peaks above the threshold in the previous instant

In previous points 3. and 4.

- counting steps with a time in between equal or bigger than 200 and 100 ms across all activity states
- counting steps with an adaptive step delay dependent on the activity state
- using 3, 10, 20 and 50 previous peaks for the calculation of each threshold value
- using a weight vector that has equal weights for each peak, bigger weights for more recent peaks and bigger weights for older peaks

Chapter 6

Methodologies for Heart Rate Monitoring

6.1 Conceptual Approach

Once activity qualification is handled in the earlier described methods, activity quantification remains unaddressed. This quantification is a translation of accelerometer values in calories spent exclusively because of physical exercise and to do so a system is needed to provide reliable values of caloric expenditure during a time period, so that a possible relation between these two variables can be assessed. As stated, in terms of energy expenditure, reliable data can come from indirect calorimetry and more specificically in open circuit indirect calorimeters, three variables are taken into account, the VO_2 or oygen consumption in ml.min^{-1}.kg^{-1}, corresponding to the difference betwen the volumes of inspired and expired oxygen , VCO_2 or carbon dioxide production in ml.min^{-1}.kg^{-1} represented by the difference between the volumes of expired and inspired carbon dioxide and UN or total urinary nitrogen in grams per day. However, since urine collection may not always be feasible, an abbreviated equation was proposed by Weir [61] to estimate caloric expenditure

$$\mathrm{EE(kcal/min)} = (3.94\,\mathrm{VO_2} + 1.1\,\mathrm{VCO_2})\,1.4 \tag{6.1}$$

Another approach that suppresses the need for measuring these both variables is to use a relation between them, a respiratory quotient (RQ) or respiratory exchange ratio. This relation can be calculated in two different environments, since the respiratory quotient corresponds to a ratio calculation of produced CO_2 by the amount of produced O_2 at the cellular level and the respiratory exchange ratio on exhaled air from the lungs ($RQ = VCO_2/VO_2$). Since heat production is accounted, it strongly depends upon the substrate metabolized and, therefore, the RQ varies with the relative quantities of carbohydrate and fat oxidized in the considered animal. To obtain the RQ for carbohydrate molecules:

$$6\,O_2 + C_6H_{12}O_6 \longrightarrow 6\,CO_2 + 6\,H_2O + 38\,ATP$$

$$RQ = \frac{VCO_2}{VO_2} = \frac{6CO_2}{6O_2} = 1.0$$

and for fatty acids:

$$23\,O_2 + C_{16}H_{32}O_2 \longrightarrow 16\,CO_2 + 16\,H_2O + 129\,ATP$$

$$RQ = \frac{VCO_2}{VO_2} = \frac{16CO_2}{23O_2} = 0.7$$

These values are then used to assign caloric values to the consumed oxygen based on standardised tables with caloric equivalents such as the ones proposed by Lusk (see table 6.1) where as an example, for a RQ of 1 there is a consumption of 5.047 kcal/L O_2, similarly to humans.

Table 6.1: Lusk table with different RQ values [62]

RQ	Calories for 1 Liter O2		Carbohydrates	Fat
	Number	Log	percentage (%)	
0.70	4.686	0.67080	0	100
0.71	4.690	0.67116	1.4	98.6
0.72	4.702	0.67231	4.8	95.2
0.73	4.714	0.67346	8.2	91.8
0.74	4.727	0.67460	11.6	88.4
0.75	4.739	0.67574	15.0	85.0
0.76	4.752	0.67688	18.4	81.6
0.77	4.764	0.67801	21.8	78.2
0.78	4.776	0.67913	25.2	74.8
0.79	4.789	0.68024	28.6	71.4
0.80	4.801	0.68136	32.0	68.0
0.81	4.813	0.68247	35.4	64.6
0.82	4.825	0.68358	38.8	61.2
0.83	4.838	0.68469	42.2	57.8
0.84	4.850	0.68578	45.6	54.4
0.85	4.863	0.68690	49.0	51.0
0.86	4.875	0.68800	52.4	47.6
0.87	4.887	0.68910	55.8	44.2
0.88	4.900	0.69010	59.2	40.8
0.89	4.912	0.69128	62.6	37.4
0.90	4.924	0.69230	66.0	34.0
0.91	4.936	0.69343	69.4	30.6
0.92	4.948	0.69450	72.8	27.2
0.93	4.960	0.69557	76.2	23.8
0.94	4.973	0.69664	79.6	20.4
0.95	4.985	0.69771	83.0	17.0
0.96	4.997	0.69878	86.4	13.6
0.97	5.010	0.69985	89.8	10.2
0.98	5.022	0.70092	93.2	6.8
0.99	5.034	0.70199	96.4	3.4
1.00	5.047	0.70307	100.0	0.0

It is considered, in similarity of other projects [63] that the majority of companion dogs have their food regime with a mixture of nutrients (carbohydrates, fat and protein) with 0.8 a proper value for the RQ with a consequent 4.8 kcal/L O_2 (according to Lusk's table [62] and a value recommended by the national research council). Nevertheless, and as stated earlier (2.1.1.2), measuring oxygen consumption is a costly and complex procedure when applied to humans and even harder when applied to animals, so there is a need to obtain some other parameter that can relate to oxygen consumption where one of the possible solutions is the heart rate. To explain this relation, it should be considered that ATP consumption in muscle cells increases as the magnitude and frequency of muscular contraction increases. If this energetic consumption is matched to aerobic respiration, oxygen consumption levels will increase and to maintain the blood of these tissues fully oxygenated, ventilation must increase in the lungs causing heart rate to rise as a way to increase the rate of blood flow to the tissues [64]. This method relies on Fick's convection equation for cardiovascular systems where it is assumed that the change in heart frequency is the main response of the animal's system to an increased demand of oxygen and the remaining terms (namely the oxygen pulse) either constant or varying in a predictable manner assuming the considered parameters' behaviour as linear [63, 64, 65].

Even though there are not many projects that explicitly report a relation between heart rate and oxygen consumption, an intense research allowed the discovery of an expression [63] where heart rate was established directly as a measure of oxygen consumption in active kennel dogs:

$$\mathrm{VO_2(ml/min/kg^{0.75})} = 0.7 \times \mathrm{HR(bpm)} - 42.65 \tag{6.2}$$

With errors below 6%, this relation applied to large healthy dogs allows for the determination of relative oxygen consumption, that when allied to the respiratory quotient determined earlier (4.8 kcal/L O_2) results on the desired energy expenditure:

$$\mathrm{EE(kcal)} = \mathrm{VO_2(ml/min/kg^{0.75})} \times \mathrm{BW}^{0.75} \times 0.0048\mathrm{(kcal/ml)} \times \Delta\mathrm{t(min)} \tag{6.3}$$

Analogously, for humans the use of MET units (2.5) allows for caloric estimation

$$\mathrm{EE(kcal)} = \mathrm{MET} \times 3.5\mathrm{(ml/min/kg)} \times \mathrm{BW} \times 0.005\mathrm{(kcal/ml)} \times \Delta\mathrm{t(min)} \tag{6.4}$$

where 3.5 ml/min/kg (or 1 MET), often characterized as the metabolic cost of resting quietly corresponds to 1 kcal/hr/kg, opposing to animal exercise where MET units of different activities are largely undocumented and demand a specific calculation (made in this case with heart rate in equation 6.2). Nonetheless, it is worth mentioning that in equation 6.3 the caloric equivalent (4.8 kcal/L) is referent to the respiratory exchange ratio, which in certain conditions such as metabolic acidosis or hyperventilation may differ from the RQ [65]. During the course of this project, an assumption is made on the equality of the respiratory quotient and respiratory exchange rate is made, even though that is not always valid.

Once the mapping of energy expenditure through heart rate is exposed, it is of value to highlight the challenge imposed by this project, which is to go one step further from referenced works and find a relation between accelerometry and heart rate making the former a proxy of the animal's energy expenditure as expressed in the diagram of figure 6.1. Such a relation will allow for the determination of a canine number of spent calories with the use of a simple accelerometer, that only evaluates body oscillation.

6.2 Technologies

6.2.1 Findster Data Channels

As in the previous chapter, a findster system with one *pet module*, *guardian* as well as the created smartphone application was used to obtain the accelerometer data.

6.2.2 Heart Rate Monitor

Two main instruments were tested regarding the monitoring of animal heart rate and given the low performance of one of them, the choice was to discard the measurements made upon the trials with that device.

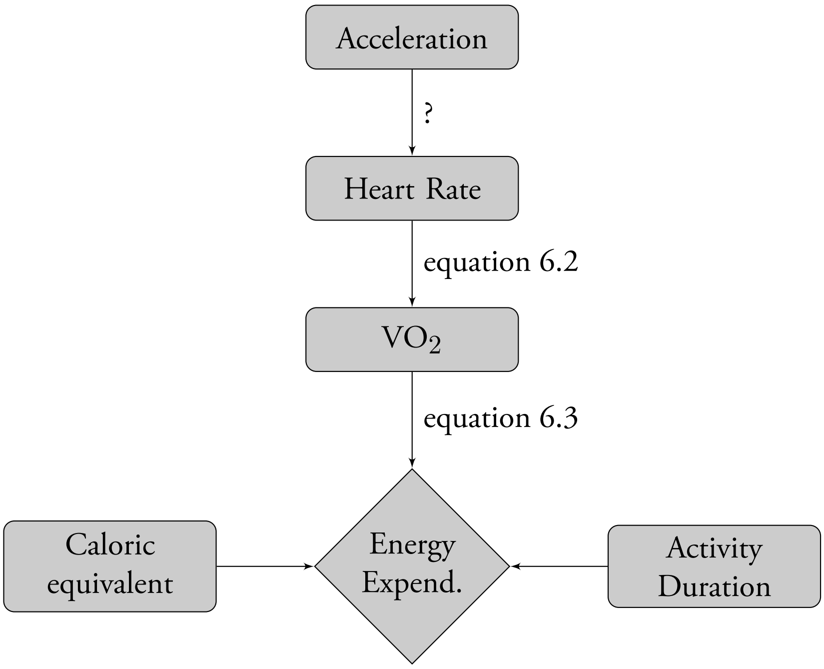


Figure 6.1: Flow diagram of relation between acceleration and energy expenditure

6.2.2.1 Polar electro CR 1632

At first, trials were made with the heart rate monitor Polar CR 1632 composed by a watch and a flexible belt possessing two electrodes that were in contact with the dog and used bluetooth to transmit the data to the watch. Even though this device is not specific for canine use, their utilization on other related projects [63, 66] had a strong influence in its choosing. The belt was placed around the animal's girth (fattest part of chest area) with the electrodes specifically below the forelegs. A high conductivity gel was applied both on the belt and on the animal for better adherence and electric conduction, but several obstacles have proven this technique to be unfeasible, namely the difficulty in obtaining reliable values without big oscillations, the short distance that had to be maintained between the watch and the belt forcing the tester to be constantly close to the animal, the lack of a writing feature on the watch that allowed for posterior analysis of the heart rate, etc.

6.2.2.2 Televet 100

After the first trials, a more robust device was used, namely the Televet 100 telemetric ECG & Holter, a veterinary equipment that provides eletrocardiogram recordings in channels with different derivations, real time recording and the possibility of storing the obtained data in an SD-card for further assessment.

6.3 Data Acquisition

6.3.1 Materials

The system was composed of a rectangular shaped device (Holter), powered by two alkaline batteries, put on the dog's back connected trough wires to four electrodes (figure 6.2) placed in regions accordingly to the coding system, where the yellow electrode is placed on the left forelimb in the sixth intercostal space, the red electrode on the right forelimb, in the fifth intercostal space, the green electrode on the left hindlimb and the black electrode on the

Figure 6.2: Televet 100 holter monitor

right hindlimb. These electrodes could be one of two types, pads or crocodile clips. In the former a need for shaving of the animal's fur for the electrodes to reach the skin as well as disinfect the area before placing them. With the crocodile clips, they were moistened with a high condutivity gel (used for ECG, EEG and defribillation) and placed on top of the fur, however, signals obtained from these clips were mostly discarded since the majority of times they slipped during the measurement and held only by the animal's hair, having no contact to the skin whatsoever. After the placement of the electrodes, the device and a SD card on the respective slot, the animal had to be bonded with a ligature capable of covering the entire equipment, as it can be seen in figure 6.3 (both 6.3a and 6.3b), while preventing the animal to touch it with its mouth. Finally the *pet module* was placed on a neck collar in the exact same way as in the previous measurements, but this time more attention was given to ensure that the device would not change its position during the measurement period.

(a) 7 year old *Labrador Retriever*

(b) 7 months old *Great Dane*

Figure 6.3: System equipped before final bandage

Concerning the measurement itself, during these trials there was a need to establish a more strict protocol (see appendix C) that whould be reproduced equally in every test for better assessment of a possible relation between heart rate and accelerometer data. Before every measurement, a stethoscope was placed on the femoral artery or chest for an auscultation at complete rest so resting heart rate could be determined beforehand. The number of heard heart beats within 15 seconds were counted and multiplied by 4 so that ground truth values could be obtained.

6.3.2 Participants

Contrasting with the participants subsection of the previous chapter, the population in this phase was composed of a group of 7 animals, however, a strict criterion was used to select them. A possible relation between accelerometry and heart rate may be dependent on several factors, but a special emphasis is given to weight, with the division of three weight groups:

- **Small size** - Canines with a weight equal or less than 10Kg
- **Medium size** - Canines with a weight between 10 and 40Kg
- **Big size** - Canines with a weight above 40Kg

The focused breeds were *Dachsund*, *Labrador Retriever* and *Great Dane* for each of the stated groups, respectively, and this choice was made with the attempt to get records of dogs with completely distinct characteristics both in the acceleration and heart rate domain. Furthermore, the selected canines had significantly different ages, which in case of a coherent relation could show a certain degree of robustness given that the algorithm could act solely based on the animal's weight.

All animals were non neutered and presented no known illnesses, heart conditions or difficulties on performing the required activities. The weight range varied from 7 Kg to 45 Kg with ages between 7 months and 7 years.

6.4 Methods

Once the data was obtained and saved upon the SD card, it was possible to verify that a direct reading upon the resultant file was impossible and so it had to be opened with the proper registered software from Televet whose key was only allowed in a specific computer located on a veterinary clinic. This meant that every recorded file had to be sent to the computer with the registered software so that it could be saved in csv format for further analysis, making this process longstanding, but more importantly not allowing for a proper inspection of the measurement process, meaning that only an *a posteriori* analysis could be made to assess the validity of the signals. This stood as a considerable barrier, translated into a big delay between each iteration. After saving the file in csv format three signals were retrieved and those where the three different derivations (green, red, and yellow electrodes), that were opened once again with Matlab.

6.4.1 Pre Processing

Considering that each measure returns 3 channels with different signals, the first step is to have a metric capable of choosing the one providing the signal in best conditions and as it is referred posteriorly, when parts of the signal are compromised, mostly by noise, the amplitude of the

ECG gets considerably bigger which means that a simple sum of each sample value allows the choosing of the channel whose signal has the least summation and probably a signal in better conditions. Moreover, a manual labelling was held during every measurement which allowed to point out the instant where the signal should be considered or in other words, every sample that is labelled with a 0 tag corresponds to an invalid point, most likely characteristic of the assembling process of the system in the animal and once this tag is changed (to 1 for example) the signal may start to be considered. The main purpose of processing the ECG signal is to obtain the heart rate during a time window and to do so the most common way is to identify key points on each cycle and assess the time interval between them for further translation in time units. One of the most common key points is the R peak, the point with maximum amplitude in the QRS complex, leading to an analysis of the distance between R peaks for further heart rate estimation. Similarly to the accelerometer raw data, the signals showed to possess a baseline that occured also because of the animal's movement, but mainly due to the thorax motion caused by respiration, resulting in displacement of both the electrodes and the monitor. Once again for its elimination a high pass digital filter was created using Matlab's "fdatool" and applied in order to remove the low frequency components such as the baseline, while this time there were no restrictions on the filter order, since this analysis is done *a posteriori*.

6.4.2 Processing

After baseline removal, the signal ought to be processed in a way that the R peak can be easily identified and for that a similar version of the Pan&Tompkins algorithm [67] was implemented with a first order differentiator applied with the function "diff" provided by Matlab that basically calculates for each sample the difference between the current value and the previous one. This operation is going to highlight quick variations such as the ones in the QRS complex and mask the slower ones. Then the signal is squared with the intent to separate clearly the R peaks from the rest of the signal and finally a moving average filter is applied to unify all peaks of a cycle into a single one, considered to be an R peak. Finally, and similarly to the prior chapter, "findpeaks" function is used to detect the resulting peaks with a few restrictions, such as "MinPeakDistance", used to assure that no consecutive peaks can be found in a certain time interval (used to avoid abnormally high heart rate values), "MinPeakHeight" established as the mean of the signal in a certain time window, used to avoid accounting peaks with a lower amplitude that can exist due to a poor performance of the average filter and finally "MinPeakProminence", used to differentiate peaks with lower distance to the baseline from R peaks.

6.4.2.1 Heart rate calculation

After ensuring that a good R peak identification was taking place, it was necessary to establish how heart rate should be calculated and two approaches were tested. At first, the difference between R peaks (ΔR) was calculated every two peaks originating several heart rate values in a time window, whose average was considered. This method calculated the often called heart rate variability (HRV) and showed abnormally big oscillations on the animal's heart frequency during a short time period. One reason for the described is a biological phenomenon called respiratory sinus arrhytmia (RSA), a heart rate variability synchronized with respiration, where ΔR is smaller during inspiration and bigger over expiration [68]. This mechanism is beneficial for the pulmonary exchange, saving energy expenditure by reducing unnecessary heartbeats during expiration. With the intent to smoothen this phenomena, following veterinarian guidelines, the second method of calculating heart rate was applied, by considering a window with a minimum duration of 6 seconds, counting the number of peaks multiplying them by 60/seconds to obtain a more trustworthy value.

6.4.2.2 Time window

So that a good insight could be taken, several time windows were tested, ranging from 6 seconds to 4 minutes and different overlaps were tested as well, in order to understand how small could a time window be for a relation between accelerometer data and heart rhythm to be found. In this case short durations were privileged giving faster response times by the system.

At first the choice for this length was made with no special care, but as several tests and iterations were done it was possible to understand that this parameter is essential to a good analysis given the fact that if the time window is too long, the animal will not be able to keep the same activity level for the entire time, particularly when running states are considered and therefore a labelling of that period is not trivial, but even worse is the fading effect of short running periods while integrating signals in these windows, not allowing for a proper inspection of the animal's oscillation in higher activity states. On the other hand if this time period is too short it is necessary to mention that while a shift in activity level is, in fact, instantaneous, a consequent change in heart rate is not, given rise to a transition phase where the heart is getting used to its new workload [69]. If the time window is too small, quick changes in activity are not accompanied by a change in heart rate, providing misleading data and with that in mind a function was applied to the raw signal were the manual labels were considered and every time there was a transition for instance, from walking to trot, the time window regarding that transition was discarded.

6.4.2.3 Accelerometry

Considering the previously established time windows, the acceleration vector (equation 2.8) during that time was calculated similarly to the activity qualifying process (5.5.3) and matched with the heart rate values. However special care was taken given the fact that sampling frequency of the devices was different (12.5 Hz in the accelerometer and 500 Hz in the holter) meaning that each time window of the heart rate signal contains 40 times more points than the same window on the accelerometer signal. The matching or alignment between the two signals was done with a reference to their ending instant, meaning that a subtraction was made between the accelerometer and holter signal's length (divided by 12.5 and 500 respectively) to obtain the difference of duration in seconds. This duration was discarded in the most prolonged signal and the reason for this alignment is held only by logistics, since there was no way to guarantee that the measure started exactly at the same time on both devices (mainly because of the envelopment done on the holter monitor) but it was possible to ensure that it ended at the same instant (it was only necessary to remove the cable from the monitor and to stop recording on the smartphone).

6.4.2.4 Accelerometry vs. heart rate

After obtaining two vectors corresponding to the values of a summed acceleration vector (or activity counts) and the heart rate, a first graphical analysis was made in order to understand if there was in fact a relationship between these two variables. Activity counts are always on the x axis, since it is the independent variable and the heart rate on the y axis as the dependent variable. At the same time a simple correlation test was applied with the Matlab provided function "corrcoef" that returns a correlation coefficient matrix, where the diagonal entries are one by convention and the off-diagonal are coefficients that point the degree of correlation between the variables where -1 is representative of a direct and negative relation, 1 a direct and positive relation and 0 meaning there is no relation at all. From this function also a p value is returned for each pair of variables. This variable is representative of the null hypothesis, *i.e.*, there is no relationship between the two variables, meaning that values close to 0 are the ones to seek.

6.4.2.5 Clustering

Resulting curves allowed for a slight identification of what seemed to be a logarithmic relation between the variables, but since a considerable dispersion was held by the resulting points, a step that could help generalizing the curves was to gather all samples from an activity state into a single point, representative of that class. For that, a relevant technique is clustering (explained in 2.4.2.5) with different algorithms tested, such as the most common k-means, k-medoids or fuzzy c-means. Clustering stands, opposing to other classification methods like decision trees or k-nearest neighbour, as the only viable method where points can be "fed" into the algorithm with no label, with different clusters or groups returned alongside with their centers. Other than applying different clustering techniques an interesting variation is to make use of the manual labelling done in every measurement, where the samples are already grouped in activity levels and the clustering technique needs only to return the center of a previously established group. Thus in every signal two tests are made, where in the first, raw samples are fed into a clustering algorithm that has to return N centers based on no classification, where N is the number of different states the animal performed in a measurement. In the second test, manually grouped points are "fed" into the same clustering that only has to return one center at a time.

6.4.2.6 Feature extraction

Considering that every clustering technique is able to group points in a N-dimensional space and in the previously described method, only two variables are used (activity counts and heart rate) an improvement may be made by extracting more features from the accelerometer signal (see 2.4.1). Therefore a set of 20 features is extracted from the accelerometer signal in each time window, both in time and frequency domain of which standard deviation, maximum acceleration value or frequency peak are some examples. Then these 20 features are added to the already existent 2 originating a set that is introduced into the clustering algorithm for further center determination.

6.4.2.7 Feature selection

Given that a big group of features was provided to the clustering algorithm, a pertinent test that can be made is whether or not each feature is relevant to classify a point as a member of one group or another, in other words, if the feature group is adequate or does not contain unwanted information that may misguide the classification. To perform this task, was used the Feature Selection Code Library (FSLib) [70, 71] provided with different methods such as filters, wrappers or embedded methods. The used group of supervised and unsupervised learning techniques like the Relief-F, Fisher or Eigenvector Centrality performed a ranking of the given features with a manual selection of a group with the 5 highest scores across all techniques and datasets. For supervised techniques the only available set of labels was used, meaning the one obtained by the manual labeling done with the mobile application. Other than simply select the most important features and considering that the results obtained with this selection were not improved it is important to note that in some data sets, groups of variables move together meaning that they measure the same behaviour of the system or in other words, redundancy is introduced into the classifier. One way to reduce this redundancy is to convert a group of variables into a single one containing all relevant information, which can be done by principal component analysis or PCA, a method that given an initial amount of variables, transforms them linearly and orders them by their capability of discriminating different classes. To apply this method, Matlab once again provides a function, ”pca”, that is able to return the new principle components and other parameters, with emphasis given to the percentage of total variance explained by each component.

6.4.2.8 Curve fitting

After obtaining each cluster center for every dataset, one point was added to the gathered group of points, correspondent to an activity count closer to 0 m/s^2 with the corresponding resting heart rate, obtained with a stethoscope. Exception made to this specific point, every other represents a cluster from a specific activity state ranging from rest to running and with this group of points, conditions were met to translate mathematically the relation between these two variables. To finally determine this relation another toolbox provided by Matlab was used, the curve fitting tool or ”cftool” that has as its inputs the x,y,z data and a list of potential curves that can be fit to the data points. From graphical visualization it was possible to limit the possibilities to the exponential, polynomial, power and custom equation curves, where different orders were tested.

Chapter 7

Results and Discussion

7.1 Canine Pedometer

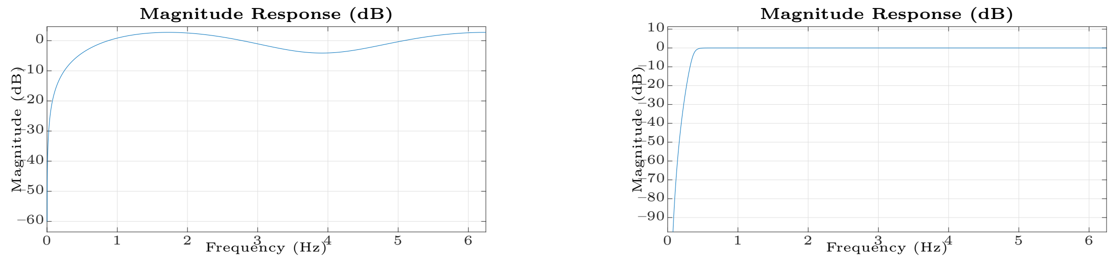
7.1.1 Step Definition

One of the most important results on the built ”pedometer” is to identify what is interpreted as a step by the system. In order to do so, and considering the device’s position on the animal’s neck, a step is accounted every time one of the two forelimbs touches the ground, meaning that for a walking stride where in fact the dog needs to move all its 4 members only the oscillation caused by front body part is registered, resulting in a registration of 2 steps. This recording changes with activity state, meaning that during walking, trotting and even during slow running, 2 steps are counted per stride, opposing to normal and intense running, where the forelimbs reach the floor almost at the same time. This was unpractical for the system to record two steps and, therefore, only one step was counted. For the static threshold that determines what peaks can be considered as steps or not, several tests were made where different animals were moved at very slow paces and the signal was analysed so that, empirically, a limit of 200 *mg* was established as the minimum acceleration of a dog step.

7.1.2 Filter Type

Choosing the filter type and order have an influence on how well the baseline of the accelerometer signal is discarded as well as slow movements performed by the *pet module* when it changes orientation. Therefore a good digital filtering process is crucial.

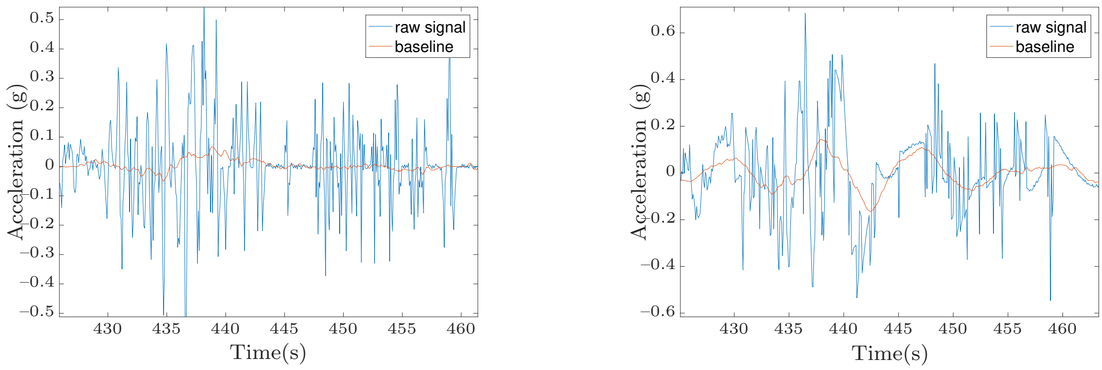
After applying the two filter types (FIR and IIR), it was possible to infer that their main purpose is achieved with the FIR filter, translating in a centered signal with an almost horizontal baseline. As it can be seen in figure 7.2a, the finite impulse response filter is able to discard more efficiently the baseline opposing to the infinite response filter (figure 7.2b) and the reason is deeply related with their frequency response, presented in figure 7.1. The IIR high-pass’s response (figure 7.1b) is highly accurate given that the cut-off frequency is 0.1Hz and the filter discards effectively all signal’s components until there, while keeping the rest (figure 5.7b) contrasting with the 5th FIR filter where the discarded signal components are closer to 1Hz and in some bands their weight is amplified (figure 7.1a), which in some scenarios can distort the signal, but in this case, that effect is not significant. Moreover, when the IIR filter was applied and tested through the smartphone application, it was possible to understand that the response had some sort of ripple response, when even after the animal stopped moving, the algorithm kept counting steps. This happens thanks to the architecture of IIR filters where previous outputs are used to determine the current output leading to small height peaks that are still considered as steps. For these reasons, it is possible to understand that for further IIR implementations on similar problems the cut-off frequency must be higher so that signal components of higher frequencies than 0.1Hz can be discarded in the frequency domain, and the baseline in the time domain. For further phases FIR filters are considered.



(a) 5th order FIR

(b) 7th order IIR

Figure 7.1: Frequency response of different filter types



(a) 5th order FIR

(b) 7th order IIR

Figure 7.2: Signal after different applying different filters

7.1.3 Filter Order

As stated earlier, two types of FIR filters with different orders were tested. The first, composed with 87 coefficients (86th order) even though it produced a more efficient frequency response, meant that more processing power was needed since 86 multiplications had to be made each time a sample arrived, contrasting with the 5th order filter whose frequency response was not so great but less processing power was needed, which in this case is critical for the good performance of the system, more specifically the calculations made in the microcontroller. Several tests were made with both filters and even though their frequency responses are visibly different the step accounting did not differ greatly, tending to sometimes be closer from reality in the lower order filter. The referred may show that even though low frequency components of the signal should be discarded for the removal of the so called baseline, they may contain important information, namely when the animal is moving in a very slow pace. Therefore it is extremely difficult to dictate an ideal cut-off frequency that centers the signal while maintaining these components and allow for a perfect signal processing, considering that this cut-off frequency exists which in fact, may not. This leads to the choice of a low order FIR filter for two main reasons, the use of less memory (allocated by the lower number of coefficients) as well as less processing power (allocated by multiplication operations between the coefficients and the signal), and also the fact that it does not exclude entirely frequency components until 0.1Hz.

7.1.4 Threshold Calculation

Analysing figure 5.8b, and resuming the applied approaches, previous step peaks were multiplied by user defined weights to build a mean sized peak whose amplitude was averaged with the baseline to produce the explicit threshold value. Still an important metric is what peaks should be considered for this calculation and as stated earlier, in one of the trials, every peak that stood above the baseline was used in this calculation which meant in most cases more steps, more threshold calculations and consequently a limit with an oscillation closely related to the one of the signal showing good results when sudden activity changes took place because the threshold is able to quickly follow the signal. On the other hand, it is possible to only consider the peaks above the threshold for its calculations, which means less calculation steps, and a much smoother line, with no big oscillations and a considerable resistance to alone peaks with extreme (high or low) amplitudes. When sudden activity changes happen, this approach seems to perform worst than the first, precisely because the threshold takes time to change its value greatly, but on the other hand when long periods of the same period take place, such as a long walk followed by small trotting episodes, the threshold shows itself to be more robust, counting the steps in a more effective way.

7.1.5 Number of Previous Peaks

When calculating each threshold value, a multiplication was made with a changeable number of peaks, and an influence of this number was assessed with the test of 3, 10, 20, 50 and 100 previous peaks. The first conclusion was that if big numbers were considered, namely 50 and 100, the threshold shows very little variance staying almost at the same value during the entire measurement, this has to do with the fact that an animal rarely keeps an activity state for more than 100 steps and considering an animal that is running, but after a while gets tired and starts to walk, this translates into a high influence on the older peaks that is cancelled by the more recent ones. Logically, if a small number of peaks is considered, the threshold has a more oscillating behaviour privileging sudden changes, meaning that an equilibrium is needed. This was achieved with the remaining tests (10 and 20) with 20 the number of previous peaks that produced more accurate results. It is important to note that the referred metrics could not be established independently since they are linked, *i.e.*, if a threshold calculation was made with all peaks above the baseline, the consequent behaviour had to be smoothened with a bigger number of previous peaks. This means that after a set of crossed tests, the described approach showed to be the one with better results.

7.1.6 Weight Vector

Concerning the weights for multiplication with the previous peaks, it was expected that bigger weights on more recent peaks would produce a threshold more closely related to the signal, privileging once again, rapid activity state changes, but what could be verified is that in most cases the influence of these weights was unsignificant and therefore a vector with equal weights was used, with the sum of all these equal to one.

7.1.7 Step Interval

In this topic, three tests were made, one with a constant interval of 100 milliseconds, where independently of the activity state if a step was detected, no other step could be identified until at least an interval of 100 ms had passed, or on the second trial after a constant delay of 200 milliseconds. The practical results showed to be concordant with theoretical assumptions since when 100 ms were applied, good accuracies were obtained for fast walking, trotting, and slow running states where in fact the interval between steps is short, but when walk, run and intense run took place, the results were not as good and the opposite happened when 200 milliseconds were applied. Therefore, an adaptive interval is in fact the best approach when on each time window, an activity state is classified and a minimum concordant step interval is defined. With this implementation the results improved although in fact, the classification system possesses a step interval dependent on the window size because when a state like walk is attributed, the interval is defined based on the previous time window. This limitation showed to be negligible, with satisfactory results, given that the length of the window was not long.

7.1.8 Step Accounting

To clearly estimate the accuracy of each applied method, the measurements made by the author were considered as well as video recordings of the animal's movement. The duration of these measurements is not long since the recording had to be done in a way that a lateral image of the dog was constantly recorded and specially when the pet was engaging in running activities, it was hard to keep track of its steps. With the video recordings providing lateral images of the animal in each instant, the estimation of the real number of steps was taken by analysing those videos several times, counting the steps and averaging those counts to establish the ground truth. Needless to say this counting process is prone to user error, which means that it is very hard to estimate exactly the real number and therefore only an approximation can be estimated. In addition to this, since the measurements were made with two independent devices (digital camera and smartphone) each of them, often, manipulated by two different people, a manual match had to be made between the video and the correspondent accelerometer signal. This process also is liable to error, since the video was not precisely stopped at the same time as the accelerometer and in those cases, some parts of the video were discarded, or if some typical parts of the signal could be identified (for instance stop states) small parts of the raw signal were cut. These factors include approximation errors on step counting, so it is expected that no algorithm is capable of retrieve a 100% accuracy since the values hold incoherence, but even so, for the choice of the best one, best accuracy values were taken into account.

Relatively to the first method, counting the number of times the signal crossed zero was a very simplistic method that served mostly as a comparison. It is possible to understand that in several cases this approach is not plausible since the curve does not reach a zero value between steps but mostly because it does so when the animal is stopped, due to the little oscillations in acceleration, ignored by other methods. The errors were in range of 6 to 45%, which is less than expected but that may have to do with the short measurement length, where it is expected that longer periods of analysis increase the error percentage. When the maximum number (between 3 axis) was considered, errors were from 11 to 48%, which is once again less than expected but these results most likely were related to the good positioning of the device, which was tightened properly, yet, if a change of the device´s orientation took place, errors would increase significantly. Concerning the threshold calculation it was possible to infer that when all peaks above the baseline are considered, there is an overestimation that leads to errors that fall in a range of 2.5 to 38% with an average error of 17% and if only the peaks above the baseline are considered, there is an overall underestimation resulting in a error range of 5 to 45% with an average error of 20%. It was also possible to verify that when only peaks above the threshold are considered for calculations the increase of its values happens easily, opposing to the decrease in threshold's height which is more difficult to happen due to the fact that if an animal is running, rising its activity levels and then stops suddenly, the algorithm will not be able to count any more peaks since they all stand below it (figure 7.3).

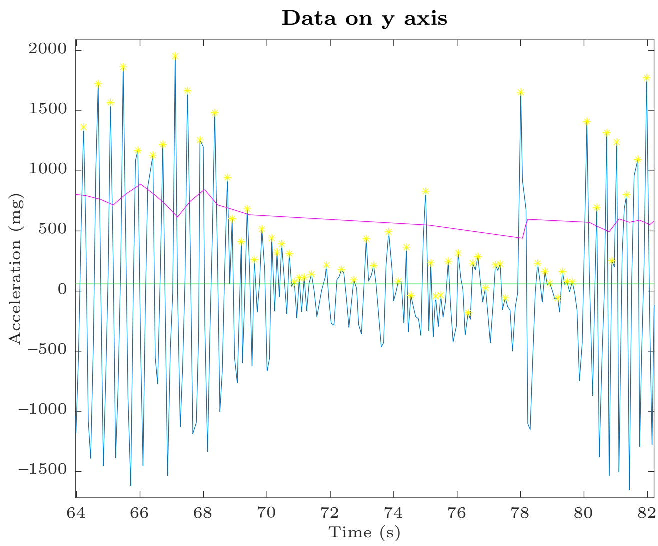


Figure 7.3: Threshold incapable of decreasing its value, ignoring a group of steps

This flaw showed to be relevant in the final classification. One of its main sources of errors, leading to the consideration of all peaks above the baseline and translating into a threshold that can increase and decrease its value in the same way. When static step intervals were applied, there was no improvement in counting performance whilst applying an adaptive step interval reduced the average error to 12%. This metric is a good way to evaluate performance since it takes consideration on how the algorithm behaves in all measurements, that recorded all different 5 activity states. Finally, as referred earlier, the weights represented no significant difference in performance so it was used a vector with 20 equal weights, with the influence of 20 previous peaks used for each threshold calculation.

7.2 Activity Monitoring

Intimately related to the analysis of the accelerometer signals, a more critical assessment was made regarding the factors that influence the dog's body oscillation, such as weight, age and gender. Breed was purposely not mentioned since the sample was not big enough to extrapolate breed-related behaviours. In order to do this analysis, a time window of 8 seconds and the pseudo-integral of the acceleration vector were used.

7.2.1 Weight

As it was predicted, dogs that possess bigger weights have normally bigger sizes translating into smaller velocity variations or acceleration, when compared to thinner (and usually smaller) dogs. Figure 7.4 exhibits the described phenomena when three dogs with different sizes (40, 25 and 7 Kg corresponding to big, medium and small size respectively) perform similar activities, showing an increase in acceleration with weight's decrease.

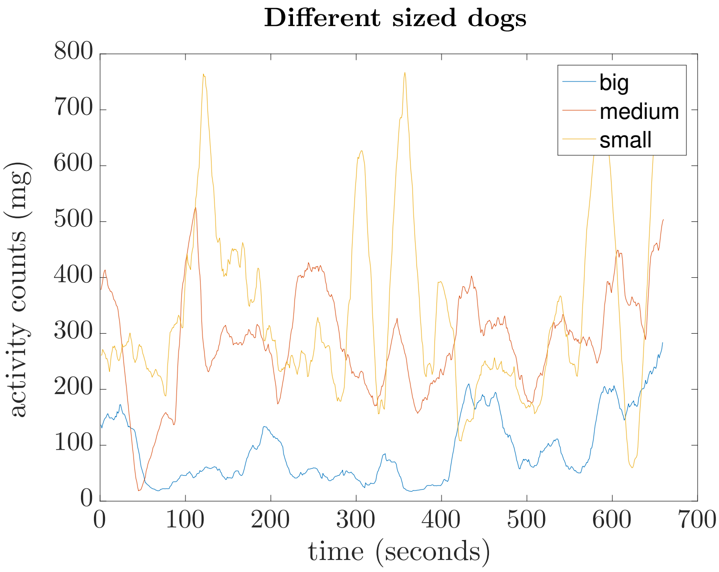


Figure 7.4: Curves of a small, medium and big sized dog (blue, red and yellow)

7.2.2 Age

Regarding age, also a theoretical hypothesis was confirmed, showing that aged dogs possess a gait with less body oscillation (or acceleration) when compared to younger ones. Figure 7.5 shows the depicted were two dogs with the same weight (25 Kg) but with different ages (6 years and 7 months corresponding to old and young respectively), performing similar activities. It is worth mentioning that this difference is increasingly visible with more intense activities in which aged dogs show more difficulty to engage, exhibiting much less oscillation than younger dogs.

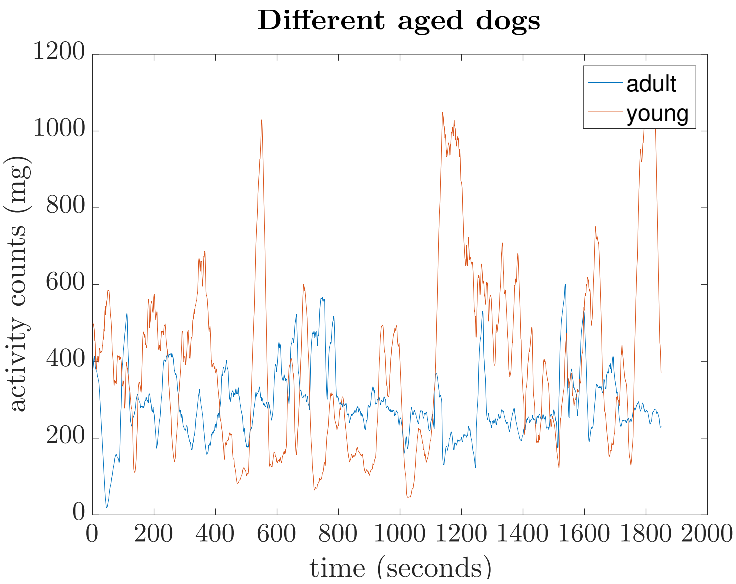


Figure 7.5: Curves of an aged and young dog

7.2.3 Gender

With regards to the gender it was possible to infer that this is, by far, the less significant factor. The figure below shows two *weimaraner* dogs (male and female) with the same age and weight performing similar activities and is possible to see (in figure 7.6) that there are no clear differences on acceleration values.

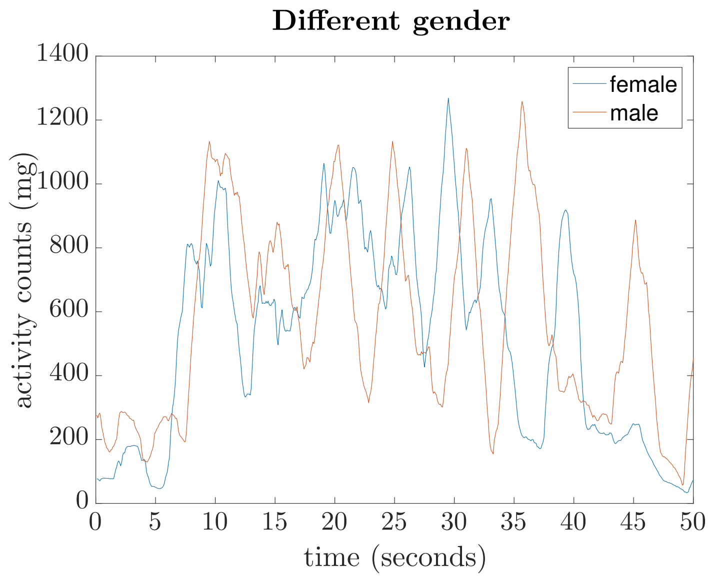


Figure 7.6: Curves of a male and female dog of the same breed

7.3 Heart Rate Monitoring

As referenced earlier, tests made with the holter monitor showed to be significantly complex, with a full cooperation by the animal which excluded pets with unpredictable behaviours and the owner presence was essential to immobilize the animal when the system was set. After some attempts to record signals with the crocodile clips it was possible to see that they were not a feasible method to extract information given the fact they are very error prone, specially when the animal performs any kind of movement (from slow walks to intense runs) since it slips through the skin holding just the animal fur and giving no information at all. Thus the use of disposable gel or pad electrodes was essential to the tests, but even with them the output was not always the intended. To minimize the electrical noise parts of the signal not only the monitor placed on the animal's back was tightened and immobilized but also the wires connecting the electrodes to the monitor were put togheter with adhesive tape to minimize movement. On top of this a special vest was made to be placed over the entire system so that it could have as little oscillations as possible, but worth to mention is the fact that these improvements on the measurement method were made based on previous experience (incrementally) meaning that the last signals contain much more valuable information than the first ones. All these upgrades tried to minimize the system oscillation but a phenomena that can never be changed is the dog's body oscillation, which in fact is the measured variable and this oscillation will cause, inevitably, movement by the holter monitor. That movement is more rough when the animal is performing high activity tasks such as intense running. From the obtained signals a visible deterioration of the ECG is noted with the increase on activity level and the analysis was impossible on states like running or intense running. This means that for the majority of tests, the animal was asked to rest, walk and trot at a speed considerably bigger than the one performed while walking, but running parts were ommitted due to this limitation offered by the holter monitor, which as it may be understood, is impossible to avoid.

7.3.1 Filter

Following the choice made in the previous section, a FIR filter was chosen over the IIR filter, but to what concerns the filter order, more meticulous metrics were needed. In other projects, high pass filters with a cut-off frequency of 25 and 40 Hz were applied [72], but considering that QRS duration in normal, healthy dogs stands between 0.05 and 0.06 seconds, faster than P-R and Q-T interval (0.06-0.13 and 0.15-0.25 seconds respectively) [73] a cut-off frequency of 20 Hz was chosen. Other than this, no restrictions were held concerning the filter's order and a transition band of 5 Hz was applied (with the stop band at 15 Hz and the pass band at 20 Hz) which reflected in a 276th order filter. However after aplying this filter to the signal, the consequent heart rate values were overestimated and that happened because of an extra peak that the filter introduced in the beginning of each time window. As it is known, filtering a signal is the same as convolving it with the filter coefficients but until all these are multiplying a signal sample there is a delay that gets bigger with a growing filter order. Previously when low order filters were considered (namely 5th order) this effect was negligible since the delay was only 4 samples long but this time, the delay was 55 times bigger and therefore an effect is clearly visible, given an example on figure 7.7.

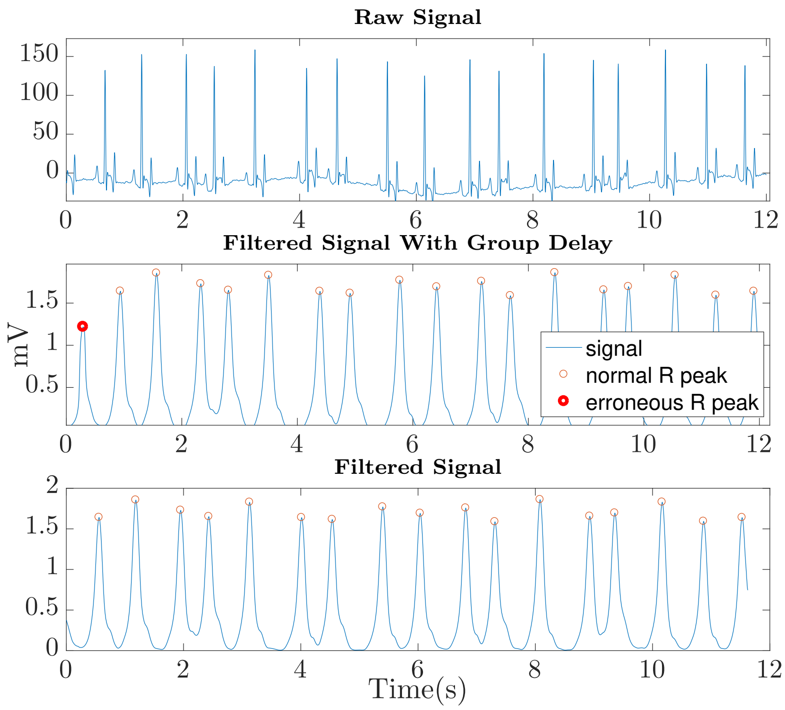


Figure 7.7: Filtered signal with and without the group delay removal

Considering that a FIR filter with a linear phase response introduces a group delay of $(N-1)/2$ samples in all frequencies, with N as the filter order, this corresponds to a delay of 138 samples, and in this particular case, 138 points are enough to create an extra peak that can be identified as an R peak and reflect on the estimation of the heart rate at a time window. To solve this issue and considering that this delay can be neglected, after filtering each time window the first $(N-1)/2$ samples are discarded so that a good peak estimation can be made. Once the filtering was done and as explained earlier a moving average filter was applied to the signal to unify the QRS complex into a single peak. In order to do so the combination that produced the best results was to first use a moving average of 50 samples and on top of that another moving average but this time with 10 samples, with the intent of strongly smooth the curve which was hard to do when only one moving average was applied. Note that after this combination of processing techniques it was no longer necessary to use either the first order derivative of the signal (applicable with "diff" function) and the squared of the signal given that the result of the first high-order filter was satisfactory.

7.3.2 R Peak Identification

After a proper filtering of the ECG signal, a clear identification and consequent counting of the R peaks within a time window was not hard while implemented the "findpeaks" function, however, some of its inputs were used in order to improve the accuracy of this R peak identification. To avoid the count of abnormaly close peaks "MinPeakDistance" was used with a minimum distance of 100 samples which corresponds to a heart rate of 300 beats per minute, the maximum heart rate achieved for alaska sled dogs when running in extreme intensity [74]. This limit assures that every heart rate can be recorded since the animals were only subjected to moderately intense exercise. Other than the distance between peaks, a minimum peak amplitude was established so that low peaks resulting from noise or other components than the QRS complex would not be accounted and for the choice of this threshold it was used the average of the ECG signal on the considered time window. At first the threshold was set as a percentage of the average, but since the signal has the majority of its points on a baseline and fewer points on the peaks, setting the minimum threshold equal to the average or even bigger produced better results. Finally to what concerns the "MinPeakProminence" input, it was possible to assess that contrasting to its application on a pre-filtered signal, improvements on the R peak identification are not significant. This has to do with the fact that with a raw ECG signal, one way to separate an R peak from a P or T wave is through the distance between the peak and baseline, but after filtering the signal that baseline is removed and only the components from QRS are evidenced.

7.3.3 Heart Rate Calculation

As referred for every signal the heart rate was calculated with two methods. On the first the time interval between every pair of R peaks was analysed and the average of those values used to determine the heart rate value for a time window. This method reflected to a great extent the synus arrythmia exposed on figure 7.8 where in a rest state and with a stable heart rate, the estimated values between each pair of R peaks differed significantly. The frequency of this change was related to the animal respiratory rhythm, which was reflected in usually overestimated values and more importantly unrealistic variations.

Therefore the second method was chosen to carry out the processing methods, where a simple counting is made on the R peaks upon a time window and a extrapolation is made considering that in a minute the rhythm is similar to the one encountered (this is done by dividing the number of peaks by the number of seconds and multiplying it by 60). As its main advantage this method aims to encompass both the inspiratory and expiratory phases of the animal's respiration, taking into account this arrhythmia when calculating the pulse. However it is worth mentioning that even with this method the final result showed abnormally big oscillations which are due to signal's parts containing noise that do not allow for a proper inspection. Several attempts were made in order to mask these outliers, where for example an heart reference was kept and if variations over 30 beats per minute occurred that time window was discarded and in the end used a method of linear interpolation to fill the missing values

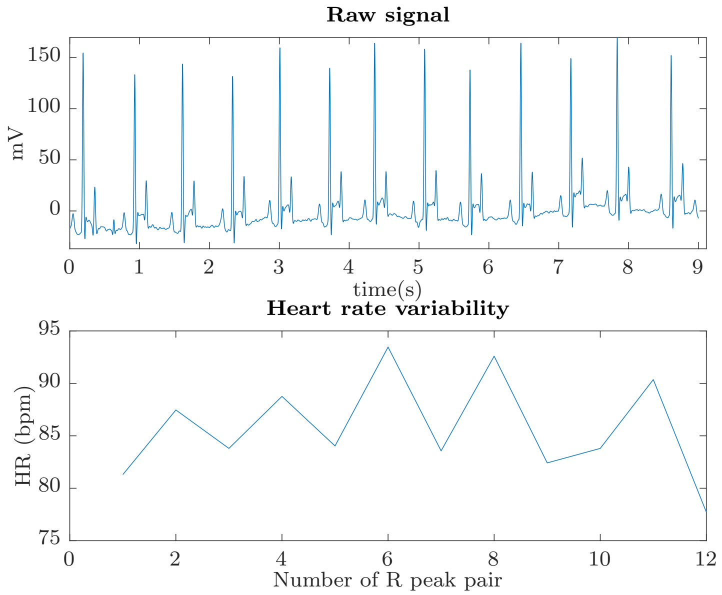


Figure 7.8: Signal portion with synus arrhythmia (below)

(with Matlab's "fillmissing" function). This attempt proved to perform poorly since discarding an entire time window, based on noise portions that sometimes occupied only part of that window removed a very significant part of the signal leaving few points to analyse and thus a simpler alternative used to smooth the heart rate curve and mask the abrupt changes over time was once more to apply a moving average filter (of 50 samples).

7.3.4 Time Window

The choice for each time window's length was based mainly on the points stated above where it must allow sufficient time for an animal's physiological response to an activity shift but not be long enough so that the animal can stay in the same activity state during that entire time. Other than these guidelines, experimentation was made with different values searching for better results. Firstly it was possible to conclude that better results came when overlapping windows were applied where correlation coefficients of 0.6 opposing to 0.3 with no overlapping windows. This may have to do with the fact that overlapping windows capture with more precision events ocurring during the measurement but the improvement is also caused by an increase on the number of points per time window, which has a positive influence when calculating correlation coefficients and p values. Regarding this method firstly an overlap of 50% and 25% was tested, but the better results were achieved when the time window advanced only one second (or half a second) in every iteration, corresponding to a step of 500 samples. This means that for instance when a time window of 6 seconds is used a non overlap method would only generate 2 points in a 12 second period but the applied overlapping window generates 6 points which is three times more (if each step's duration is half of a second the increase is six times bigger). To what concerns the time length, different values were tested ranging from 6 seconds to 4 minutes and with respect to the R and P values (correlation coeficient and null hypothesis coeficient respectively) a constant increase on the window's length reflected a bigger correlation but at the same time a bigger p value which represents a bigger probability of the null hypothesis (the variables are not related in any way). When windows were bigger than 1 minute p values were above the default limit of 0.005 indicating that the length should be smaller but considering that short time periods are more feasible to be performed by the animal, periods of 12 and 6 seconds were chosen for the time windows and R coeficient values of 0.65 and 0 for p value were obtained which can reflect a clear relation between the variables and eliminates the scenario where acceleration poses no effect on heart rate. Moreover, when compared to [36], which stated "10-second example duration because we felt that it provided sufficient time to capture several repetitions of the (repetitive) motions involved", the choice seems adequate.

7.3.5 Clustering

Regarding the applied clustering techniques their main purpose was to reduce an entire group of points, related to an activity class such as rest to a single point or center. These different techniques applied such as k-means, k-medoids, fuzzy c-means, subtractive and hierarchical clustering. Instead of visually assessing the results of all techniques which is cumbersome and error prone, accuracy was calculated for every method based on the *a priori* manual labelling, *i.e.*, the labels attributed after each clustering technique were compared to the ones attributed by the user and a 100% of accuracy meant equal classification. However it is worth mentioning that clustering, as an unsupervised learning technique only physically separates groups of points and does not have the knowledge to attribute them a class (which is in this case 1 to rest, 2 to walk, etc) and to avoid this problem, for every technique, two center groups were compared.

Considering one dataset of labelled data that is fed into a clustering algorithm, results in, for example three cluster centers A, B and C, corresponding to states of rest, walk and trot. When unlabelled data is introduced into a different clustering technique three new unlabelled centers D, E and F are returned. Then the distances from center D to A, D to B and D to C are calculated and the minor chosen to assign a class to cluster D. This process was repeated for E and F.

A simple first analysis led to the exclusion of hierarchical and subtractive techniques given that they produced results with low accuracy, leading to the use of k-means, k-medoids and fuzzy c-means. Surprisingly, the biggest differences in accuracy were caused by the change on the distance metric instead of the technique itself. In the majority of cases, between the used techniques occurred differences of 1% or less in accuracy, but when cityblock or manhattan distance was used a bigger decrease was noted, leading to the choice of the euclidean distance. Relatively to the technique given that no significant changes were felt, k-means was the selected mainly due to its bigger popularity among classification problems and to the better knowledge of its algorithm. Concerning the different center location when applying clustering to unlabelled or previously labelled data, which is the main reason for this step, clear diferences are visible, as in figure 7.9. The differences on x and y coordinates are not sistematic but a tendency may be found for the raw labelling to estimate centers with a bigger x coordinate (or activity counts) meaning that the centers are dislocated to the right when compared to the clustering with previously labelled data (figure 7.9 as an example).

Other than that, a systematic phenomena is that decreased activity states usually reproduce closer centers. This can be explained on what was already mentioned that during rest the animal performs less movement, the ECG signal is more reliable as well as heart rate's extraction, making these groups possess less outliers and allowing a better partition and center estimation. With the increase of movement, in walking states, even though the animal's gait is controlled more outliers can be seen reflecting on a bigger diference between centers of

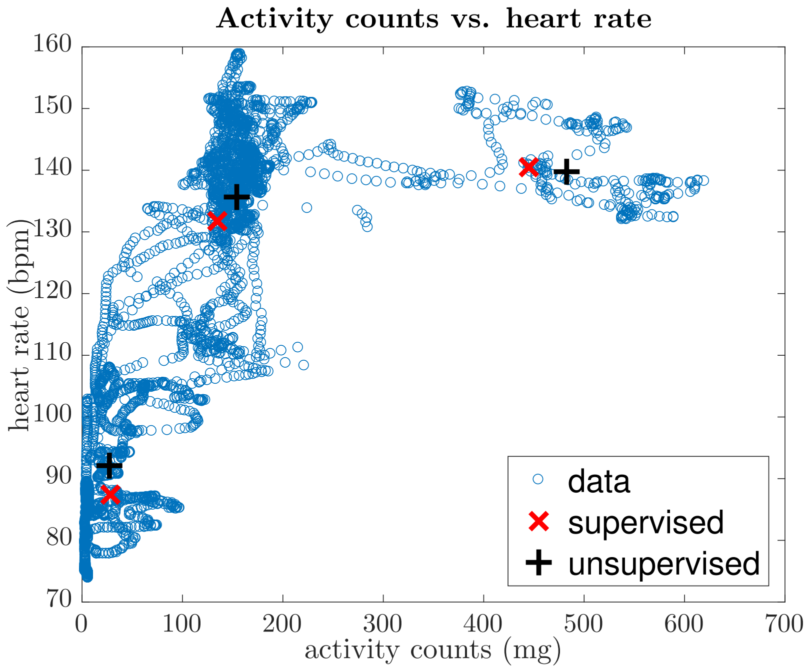


Figure 7.9: Features' cluster centers with and without manual labelling (super and unsupervised)

distinct techniques and from trot until running, where less control can be held both on the animal's gait, velocity but mostly on the holter's oscillation, bigger parts of the signal contain significant noise that do not allow a proper heart rate estimation while inserting values that may not be reliable, giving centers that are clearly apart as it can be seen on the figure above. In order to fit a curve to those points, which is the final goal of the project, a choice has to be made on what points to fit and this choice cannot be made just by picking the centers resultant from the previously labelled data given that this labelling is not perfect and it is done manually. Therefore the centers obtained by both methods were kept for further curve fitting.

7.3.6 Feature Extraction

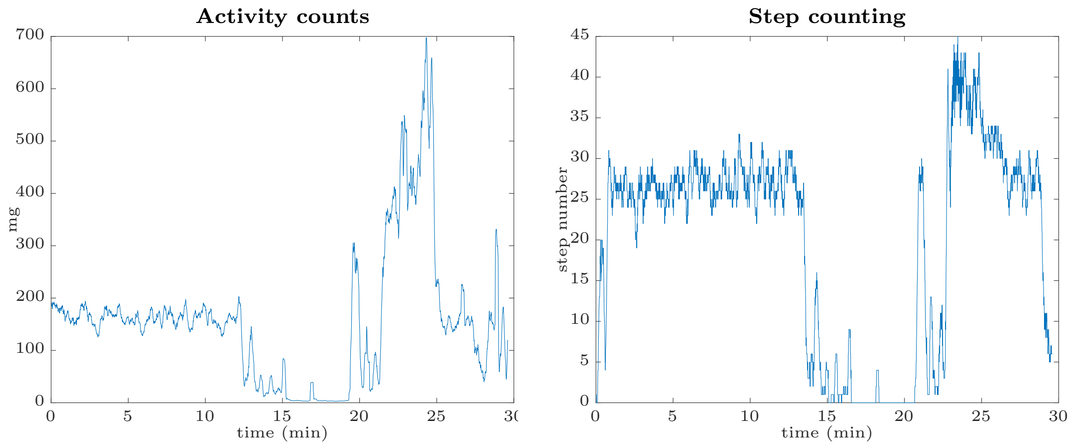
After applying the mentioned clustering techniques in order to determine centers in a 2 dimensional data space (activity counts as x and heart rate as y) a possible way to improve the determination of these centers was to provide more information to the algorithm with a group of features added to the already existent and turning the data space into a 22 dimensional one. This group is composed of the maximum SVM in a time window, the minimum, mean, variance, standard deviation (all of these of the acceleration vector which encompasses information of all axis individually), the distance of every point to the mean (referred as mean average distance - mad), the skewness and kurtosis obtained within the time domain.

In the frequency domain, the entropy, window spectrum centroid's frequency and its spreadness level calculated by checking the similarity between the obtained spectrum and a rectangular one. This feature is related to the fact that if the window's spectrum is highly spread, all frequencies are present leading to the conclusion that the signal was taken in most part by noise not allowing its analysis. It is also taken the bigger harmonics (or main frequency) as well as its intensity (y coordinate), the area under the spectrum curve and its entropy. Finally a cepstral spectrum is calculated [37] by doing the logarithm on top of the signal's fast fourier transform as well as its correspondent spectrum, the area under its curve and its entropy. Adding to the 18 described features, it is extracted the number of steps, obtained by applying the algorithm created in the previous chapter and the animal's activity state on the previous time window, which can be a good indicative considering the fact that most of the times the activity transitions occur gradually, for example, transitioning from rest to walk and from walk to trot. After performing clustering in the new 22 dimensional space, the results show no improvement and there was a slight decrease on accuracy noted across the majority of signals. This allows for the interpretation that some of the introduced features include either redundant or misleading information, translating into a worse classification by the algorithm and producing centers that, probably are not representative of the considered activity state. It is worth mentioning that any clustering algorithm when analysing instances with 22 features returns centers with 22 coordinates and therefore for them to be projected in the previously given 2D space only the first two coordinates are considered (given that the two first features are activity counts and heart rate, by this order only). This systematic decrease in performance led to the next sub section, but it may perhaps be observed that all features are extracted from the accelerometer signal and none from the electrocardiogram. This is due to the fact that in most cases, ECG's features are related to peaks' positive and negative amplitudes or areas, time-interval durations of not only the QRS complex but also of the P and S wave [75]. However, as said earlier, the holter's signal is affected by a lot of noise not allowing for a clear determination of the P and S wave neither their amplitudes or time duration and a proof of that is the fact that processing techniques had to be made in order to extract the R peak, the easiest point of extraction in an electrocardiogram.

7.3.7 Feature Selection

As mentioned, Feature Selection Code Library (FSLib) was used by applying a group of 10 different techniques into each signal to assess which features were more discriminant or in other words, the ones who could classify best the different activity states by their own. To do so, the techniques needed features as the ones used in previous clustering and a label for each instance, This label group was the manually attributed one. The applied techniques were Inf-FS (infinite feature selection), EC-FS (eigenvector centrality), Relief-F, mutiinffs (mutual informaion), Laplacian score, MCFS (multi cluster), generalized fisher score, regularized discriminative (UDFS), pairwise correlation (cfs) and kernel learning for local learning-based clustering (llcfs). All 10 techniques were applied to the 22 feature set possessed in a signal where the group with 5 highly ranked ones was selected. This group was composed of the maximum acceleration, variance, spectrum rectangular area, step number and previous activity level. Once again a new feature set, this time with 7 different variables (5 selected and 2 already used) was introduced into clustering with no improved results. The lack of improvement can be explained if it is taken into account that the selected features are not systematically the ones with the highest ranking across all feature selection techniques meaning either the techniques are not robust to select the relevant information or the features themselves are not reliable. The first hypothesis is discarded given that distinct techniques are applied and therefore it is unlikely that all of them are not effective enough. To select the most important features, it was counted the ones that stood on the top 5 ranking of all techniques with the more frequent ones selected, but even doing so an evident discrepancy was found which led to the exclusion of the ones that had theoretically less capability of finding patterns in the data, such as the intensity of the first harmonic. Considering the selected features, it is easy to understand that step counting, mean acceleration and previous activity level may be discriminant of the activity level given that in intense activity states the number of steps and the mean acceleration are bigger and the previous activity level corresponds normally to rest related states.

Nevertheless as said there were no improvements and an attempt to explain this behaviour was held by a linear transformation of the 20 feature set applicable by principal component analysis ("pca" function in Matlab). The main idea was to perform this transformation on the 20 extracted features, extract the most relevant ones, add them to activity counts and heart rate, apply clustering and check for accuracy, however an interesting output provided by this function was analysed at the same time, named percentage of total variance explained by each component. After applying pca, it was interesting to notice that 95% or more of the total variance was explained by only one variable, which allows for the conclusion that extracted variables are secondary and contain redundant information, which is the same as saying that they are variants of basic features. That happens since pca performs linear transformation of the features returning a new set of orthogonal ones and if only one of those directions contains almost all variance of the signal this means that variables change similarly, as the figure 7.10 can illustrate, and this redundancy is introduced into clustering resulting in no improvement on performance.



(a) activity counts

(b) steps (feature number 19)

Figure 7.10: Different features expressing the same features only on different scales

Considering that all features are extracted from the same signal (accelerometer) it is possible to assess that the same variations are evaluated only with differences in scale. In sum, it is possible to declare that feature extraction in this particular case, with this particular feature set is not effective and performing clustering with only 2 features produces the best results given that no redundant or misleading information is introduced. Therefore the centers obtained by those scenarios were the ones considered for the remaining steps.

7.3.8 Curve Fitting

Once centers were calculated, Matlab's curve fitting toolbox was used in order to obtain the best fit, but before doing so a mathematical, theoretical approach was performed. At first a key point obtained by animal's auscultation (6.3.1) was considered, representative of a complete rest period which allowed a bare notion of the interval in which heart rate values should be contained. Results suggest that heart rate values evolved logarithmically, *i.e.*, the heart response of an activity shift such as, from rest to walk is much more significant than, for example, from walk to trot, evolving this curve until a horizontal asymptote. Worth mentioning is the fact that this analysis is not possible in extremely high activity states such as intense run given the provided equipments' restrictions.

To determine the best fit, the coefficient of determination or r-squared is assessed considering that it returns the percentage of heart rate variation explained by the model, yet, this measure cannot determine whether the coefficient estimates and predictions are biased, and to avoid it, residual plots as well as the sum of squared error (SSE) are analysed too. Logically, high r-squared and low SSE values are privileged.

7.3.8.1 Model type

Similarly to what was found in humans, "using a single linear regression to estimate EE cannot obtain satisfactory accuracy when physical activity intensity levels range from sedentary to vigorous" [35] and therefore other models were tested:

- **Power Curve** - In this option, the toolbox fits a curve with two possible equations, $y = ax^b$ or $y = ax^b + c$ corresponding a,b, and c to the free parameters, x to activity counts as the independent variable and y or heart rate as dependent.

  Values of 0.83 and 520 are achieved for r-squared and SSE respectively, but with the introduction of the c coefficient an improvement is reached to 0.85 and 453.

- **Polynomial** - Even though r-squared values closer to 0.8 were obtained, models with equations like $y = ax^2 + bx + c$ with 2 or more degrees, contain a negative concativity which meant, at a certain point, a decrease of heart rate caused by an increase in activity, which is false in every case.

- **Exponential** - This model type can possess equations like $y = ae^{bx}$ or $y = ae^{bx} + ce^{dx}$ where the first has a number of terms equal to 1, the second to 2 with the second capable of producing satisfactory results with corresponding values of 0.87 and 399 for r-squared and SSE. However this equation seems too complex and there is a bigger possibility to be overly adjusted to the obtained data.

- **Custom equation** - Observing the shape of previously fitted curves a manual approach was attempted with the main goal of mimetize a step response of a system over time.

  In first order systems a step response is characterized by $\frac{b}{a}(1-e^{at})$ where $a > 0$, $T = \frac{1}{a}$ or time constant referent to the time needed for the system to reach the asymptote $\frac{b}{a}$ and in order to mimetize that behaviour in this particular case, time was equated to activity counts and the following similar equation was introduced into the Matlab's toolbox:

  $$y = -a\, e^{-\frac{x}{b}} + c \tag{7.1}$$

  which produced results with r-squared values up to 0.92, SSE of 400 and residuals that do not exceeded 10 beats per minute, and with the majority included in an interval of [-5,+5]. At first the expression was not converging to a reasonable result, but that happened because the starting point of some parameters, more specifically the b parameter had to be set manually for a value closer to the expected so that it could converge after a certain number of iterations.

  Even so, in one particular signal, before obtaining the coefficient values by the toolbox a manual approach was attempted taken into account both the resting and maximum heart rate obtained by a stethoscope. In order to illustrate properly the referred a medium sized dog was measured, having a resting heart rate of 86 beats per minute (after averaging some auscultations) and a maximum heart rate (on mediumly intense exercise) of 150 beats per minute, this meant an horizontal assymptote, or in other words, a c value of 150 correspondent to an infinite x value. If on other hand the animal is stopped and x is closer to 0, a must take the value -64 so that heart rate can be 86 bpm and finally, b is the parameter that sets the "speed" in which the curve reaches 150 bpm. By trial and error the value 100 was attributed for testing purposes leading to the expression

$$\mathrm{HR} = -64e^{\frac{-\text{activity}}{100}} + 150 \tag{7.2}$$

After introducing the same equation on the toolbox but with undetermined coefficients the results were a $= -55.67$, b $= 101.5$ and c $= 137.2$ representing closer values and a similar model prediction.

If second order systems were to be considered, the behaviour would look similar to an overdamped step response, but since the final equation does not differ much from the one obtained by the exponential approach provided by the toolbox neither its complexity is reduced, mimetizing this type of systems was not tested.

7.3.8.2 Weights

Taking into account the already stated, points provided to the model descendant of lower activity states are (theoretically) more reliable given that less noise obscures the ECG signal when opposed to states where moderate/intense animal movement takes place and in an attempt to reflect this influence on the model fitting, a weight vector is introduced alongside the points' coordinates (one weight value for each data point).

To determine the number of times the weight of a resting point should be bigger than a trot or run point, a signal to noise ratio (snr) analysis was made on each signal with the function "snr" provided by Matlab. This function takes two inputs, the signal and noise but since it is not possible to extract exclusively the noise a simple approach is used by taking into account the pre and post filtered signals, where the filtered signal is the first input and its subtraction to the pre-filtered signal stands as the second input. This function is applied to all windows and the resultant vector is assessed visually allowing for the determination of how many times the snr is bigger during rest when compared to trot or running. This relation reflected on the establishment of the weights. As an example if rest snr is 3 times bigger than run snr, points of the former will have a weight of 3 and points of the latter a weight of 1. The introduction of these weights resulted in r-squared improvements, from 80% to 90% but in an increase in SSE, from 400 to 660, which happens because the model tends to approximate more the first points regardless of how far it is of the latest ones, reflecting into a lower c parameter or horizontal asymptote which in practical terms could mean a lower maximum animal's heart rate as it can be seen in figure 7.11, which presents a screenshot of the "cftool".

Since for some cases the horizontal asymptote was already smaller when compared to the maximum heart rate obtained by auscultation, the application of weights was not held in the final model.

7.3.9 Model Generalization

As stated in subsection 6.3.2 the described steps were performed on a sample of dogs with different characteristics with a special emphasis given to weight groups. Three weight and size groups were considered and results show that the model obtained by the custom equation 7.1 is better across all subjects.

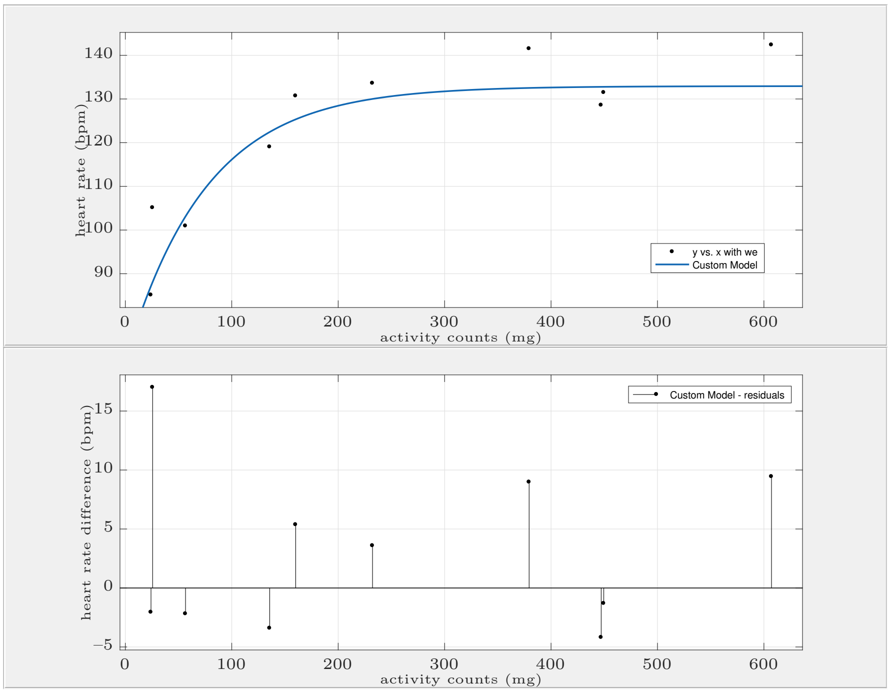


Figure 7.11: Graphs generated by the "cftool"

7.3.9.1 Free Parameters

In $y = -a\, e^{-x/b} + c$ there are three free parameters ("a", "b" and "c") where findings suggest than those are characteristic of a tested animal:

- Even though the parameter "a" is not, directly, representative of any information, "c-a" provides the starting point of a specific model's curve, which may correspond to the dog's heart rate while at rest.
- "c" on the other hand, as an horizontal asymptote gives directly the value to which a dog's heart rate tend to in mediumly intense physical activities.
- "b" also gives information on the tested animal, more specifically the capability of a given heart to resist an increase in activity intensity with little increase on heart rate. This type of "inertia" can be interpreted as the heart endurance, where bigger "b" values correspond to dogs in better physical shape when compared to ones whose model has lower "b" values. This happens because when an increase in activity counts takes place, models with higher "b" will have a smaller increase in heart rate when compared to models with lower "b", showing a smoother behaviour as it can be seen in figure 7.12, on the *Great Dane* curve.

7.3.9.2 Weight Groups

The main attempt of this project was to build three models based on the animal weight and size, assuming that regardless of other factors, such as age and breed, dogs with smaller weight had bigger heart rate values and consequently, bigger "c" and smaller "a" parameters, corresponding to higher resting heart rate and higher limits in mediumly intense activities (at rest). Yet, it was possible to infer that that it is not the case, which is represented on the right graph of figure 7.12 and in agreement with [76, 77] that weight is not enough to characterize the animal's heart rate profile nor it's variability.

Analysing figure 7.12 in more detail, the main differences between theoretical and obtained results are held within the *Great Dane*'s curve, which fluctuates between the other two curves, opposing to our hypothesis, when it should be below them. The answer for this issue may be related with the animal's age, since the *Great Dane* is still on his puppy stage, with only 7 months old in contrast with both the *Dachsund* and *Labrador*, with 5 and 6 years respectively, representative of a clear adult stage.

Another aspect worth of analysis is the diferent "b" parameters of the red and blue curve, causing an overposition of both curves in contrast to what was predicted. This phenomena may be described by the physical condition of the *Labrador* which is overweight, with 5 Kg above its ideal weight. This state may be the cause for a curve that reaches the assymptote with a smaller increase in activity, given that the animal presents more difficulty when performing exercise. In its ideal weight, it is expected that the behaviour of this animal's model resembles the one in the left graph of figure 7.12.

Lastly, another important note must be held with the horizontal asymptote of the *Dachsund* animal, whose difference from the other curve's limits may be exacerbated. Theoretically this difference should not be as big and one reason for that could be related with the measurement process, where an external factor, namely the arrival of the animal's owner raised significantly the animal's heart rate originating peak values that were interpreted as points of mediumly intense activity where in fact they were not. This limitation related to the animal's temperament (explained in the further chapter) was significant, with the expected behaviour characteristic of a similar curve but with a smaller horizontal asymptote.

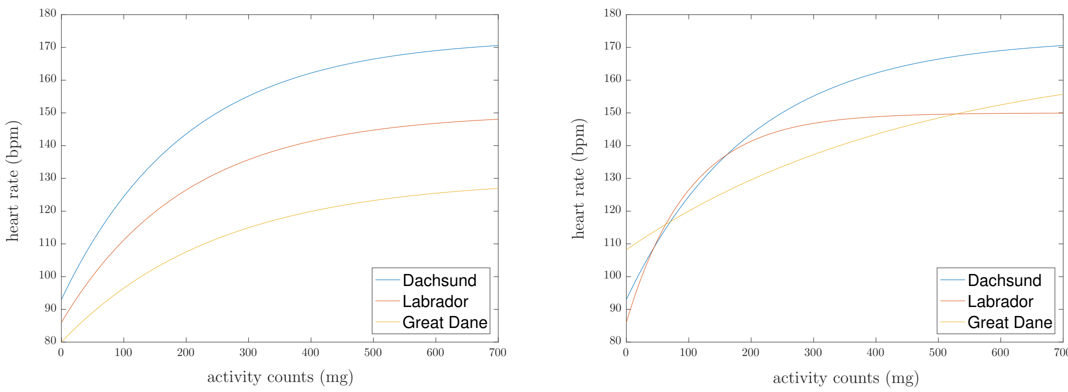


Figure 7.12: Predicted models (left) vs. Obtained models (right)

7.3.9.3 Small Weight Group

Given that with such a small sample it was not possible to extrapolate three distinct models to the whole canine pet population, a more specific approach was made in this second test phase where only dogs belonging to the small weight group were analysed. This was done with an attempt to restrict the variation in body weight and assess if any other factor, such as age or breed is more discriminant when defining a relation between acceleration and heart rate (and consequently energy expenditure). Firstly and observing figure 7.13, it is worth to mention that the blue curve (Dog1 model) is referent to the *Dachsund* model of figure 7.12, where the horizontal asymptote is an exaggerated value. Taking that into consideration, the most evident fact is that all curves tend to a similar value, even with canines with completely different ages (Dog3 and Dog4 with a difference of 13 years). Moreover the blue and orange curves, of two animals with the same weight show a very similar shape with a bigger age responsible for the decrease in the overall heart rate values. To what concerns the resting heart rate, it is possible to see that there is a bigger variation with no systematic variation when compared to weight, age or breed.

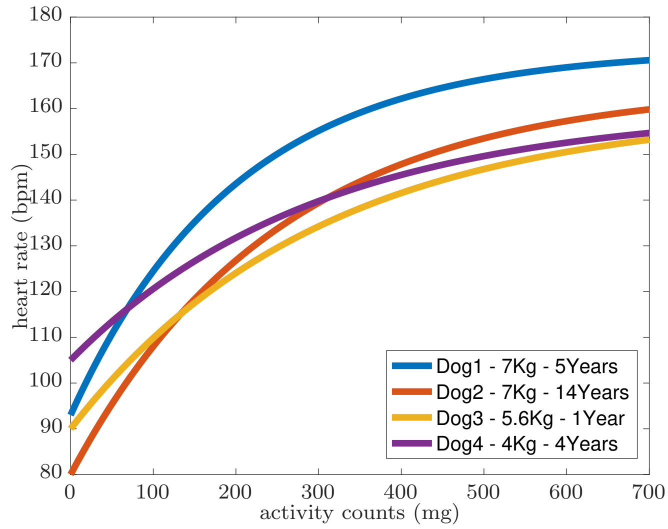


Figure 7.13: Models obtained from dogs with small weight and size

With these results, suggesting a similar logarithmic behaviour tending to a value close to 160 beats per minute, a final model was created by averaging the obtained models and unifying them into one.

$$y = -71.82 \times e^{-\frac{x}{251.4}} + 160.7 \tag{7.3}$$

Equation 7.3 is proposed by this project as a generalized model of obtaining heart rate (and consequently energy expenditure) from acceleration, for dogs of small weight and size.

Chapter 8

Conclusions

From a general perspective, the biggest challenge on both parts of the project was to create solutions that worked appropriately across different canines. As expected, age, weight, breed and even individual variations are important factors that cause differences in dog's body oscillation and heart rate profiles.

Regarding the creation and implementation of the canine pedometer, it is of value to mention that it presents reasonable results across factors as the ones determined before. Even so, it was possible to verify a more effective analysis on dogs belonging to the medium and large weight groups, given that with smaller dogs more steps are given per unit of time and the algorithm presents more difficulty in distinguishing spontaneous body oscillation from a step, given that the correspondent signals are very similar.

Regarding activity quantification and the novel multi-proxy method of obtaining energy expenditure from acceleration, more conclusions can be withdrawn. A first division based on weight groups showed to be inappropriate, given that the model curves of breeds with different sizes do not show a clear distinction, with big sized dogs having higher heart rates than small sized ones and even occuring an overposition of these curves, opposed to the expected results. These results suggest that weight is not a factor explanatory enough to subdivide animals in static intervals and other factors must be taken into account for a proper evaluation.

Focusing on one weight group, the results show more coherence with much more resemblant curves, tending to a similar value of heart rate. In this particular case, factors like age and physical shape (expressed by the "b" term on equation 7.1) see their influence be diminished (even though it is far from null) allowing for a generalized model. Beyond the already described factors, the visible variation on resting heart rate by different animals suggest that not only there is variability inter species, but also intra species, where animals from the same breed can have different heart rate values across several intensity levels of physical activity.

8.1 Limitations

To what concerns the pedometer it seems not to be limitations on the created algorithm since a minimum limit was established for a step's acceleration and a maximum/minimum time between two consecutive steps (respectively to identify resting periods and to avoid counting extra steps) was applied to the vast majority of dogs. In contrast, the activity qualification based on the proposed multi-proxy method presents a group of limitations that must be referred:

- Gender - As stated in [15] energy expenditure is proportional to the number of metabolically active cells in the body with the majority of them lean body tissue which in humans is different in relation to gender (women have lower percentages relative to men). However, this gender difference in energy expenditure is not evident making the created model applicable to both male and female dogs.

- Temperature - If a subject is in their thermoneutral zone, defined as the ambient temperature which above or below, the resting energy expenditure begins to rise in order to maintain body temperature (18 to 25 degrees Celsius in most dogs) [15], the effects of a change in temperature may be negligible, however if other conditions are met, temperature can exert a large influence on caloric estimation. This meant special care for every measurement, that was made during late afternoon period with not too much heat (possibly causing increased energy losses via sweating or panting) nor too much cold (causing energy losses by shivering and non-shivering thermogenesis).
  Even though for each degree Celsius that is increased in core temperature there is a consequent 13% increase in resting energy expenditure, it is necessary to take into account that the presented equation only aims for the quantification of caloric spending caused exclusively by engagement in physical activity, *i.e.*, if a dog is running in an extremely cold environment its total energy expenditure may be higher than the one calculated, but the presented value gives a notion in how much of that expenditure was caused by exercise.

- Alimentation - The proposed method relies on indirect calorimetry principles that use heat loss as a way to measure energy spending, but it is important to note that during and after a meal, the total energy expenditure increases, mainly in two phases, the first due to the stimulus of the palatability of a meal (also called cephalic phase) and the second with the digestion, absorption and nutrients' storing. The duration of this second phase is not constant across animals, depending on the meal's size as well as the feeding frequency, meaning that every measurement had to be made within a period of at least 3 hours after dog's feeding.

- Temperament - Heart rate variability could be caused by quick sympathetic activation, for example, due to excitement or stress [63] and although this was minimized, every time the animal understood that a walking activity was taking place short increases in heart rate took place which affected the measurements resulting most likely in outliers. This effect was however minimized with the application of clustering techniques that took several points forming a cluster and determined a center in the data space, reducing the influence of the stated outliers.

- Age - Heart rate can vary within a specific breed in different age sectors, with higher values in dogs less than one year old, lower in young adult dogs and increasing with age in dogs over four years old [77]. Given the small group of participants during the second phase of measurements, which accounted mainly for differences in size/body mass, different ages and intra breed analysis is a topic to study in future work, with the presented model applicable mainly for small sized dogs.

- Others - A number of other factors can affect heart rate, like exercise regimen, athletic fitness, stage of oestrous cycle, sleep, activity, medication and pain [77]. Some of them were avoided during measurements but there might have been effects caused by unknown conditions.

- VO_2 vs. HR regression - Possibly the most evident limitation of the proposed method is the fact that a single regression is used in order to obtain oxygen consumption through heart rate. The used equation (6.2) is valid for dogs weighing around 30 kg with a maximum heart rate of 190 bpm [63], but as stated in the article future work with different animals must be made to refine the predictions over all companion dogs.

- Different accelerometers - The created models, placing heart rate as a result of activity counts are very dependent upon the used accelerometer, with different sampling rates and signal filtering. Yet, if an adaptation of this model must be made to that specific accelerometer, a simple validation test must be made to prove the existence of a linear relation between NXP's MMA8652FC and any other accelerometer activity counts, similarly to [78].

8.2 Advantages and Applications

To compare extensively the proposed method of energy expenditure calculation with other methods descendant of direct/indirect calorimetry is beyond the project's scope, however, it may be relevant to note that it represents a much easier alternative when compared to existing methods, requiring only the use of an accelerometer placed on a neck collar, a device that is cost efficient when compared to expensive direct calorimeters, easy to assemble when faced against portable calorimeters that need to be bound to the animal in specific laboratory conditions and also capable of providing long time recordings in both modes of Findster's system. In Live Tracking mode given that the system communicates data every 10 seconds the battery is capable of lasting up to 7 days and in Offline Tracking communication happens only when the user asks to, allowing the battery to last up to one month.

Lastly, this method stands as a non-invasive technique, where no discomfort is caused to the animal when compared to, for instance the heart rate method where a bandage needs to be set around the animal's chest. Also an advantage of this method is related with the fact that during measurements, the conditions experienced by the animals which were subject of research matched almost entirely the ones from their real life activities, with the performance of slow and fast paced walks without a leash (given that it could interfere on the oscillation profile) followed by periods of rest.

Given equations such as the ones provided by this work allow, generally for a better monitoring of a companion animal's spent energy, from reports on metabolic rate of different free ranging activities or behaviours, environments (physical and social) to a more accurate match in calculation of energetic requirements and energy intake from food, opening space to a better understanding of each breed within a specific set of conditions [64]. In addition to all of this, weight management programs can be made based on this exercise caloric estimation where the owner can have a better insight about the amount of exercise practiced by his or her pet and adjust it so that it can meet the animal's ideal requirements and culminate in a ideal weight management. Yet, it is crucial to note (as it has been made throughout the project) that physical activity only stands as part of the solution for weight loss (or gain), with the dietary regime also a cornerstone for this issue.

8.3 Future work

As possible future work may be the incorporation of a bigger group of animals both in small and non small weight gorups, tested under the same conditions to support on a larger scale the proposed models. With more subjects it may be possible to infer other factors affecting energy expenditure, such as age, medication, different diseases, etc. Moreover, it is important that more tests are made regarding the relation between oxygen consumption and heart rate

(VO_2 vs. HR) so that an even more accurate relation can be held in the final model of heart rate vs. activity counts. Finally there is a huge opportunity to follow this project if there is some equipment capable of recording dog's heart rate at extremely intense activities. With that analysis, models can encompass all possible activity states by the animal and extend the x-axis of images such as the ones in figures 7.13 and 7.12. Considering the main target audience of Findster Technologies, the majority of canine population do not engage in extremely tiring activities, privileging moderate or brisk walks, but if markets such as hunting or races should be considered, future work can and must be held, with this work as a good starting point.

Bibliography

[1] The Free Dictionary. Obesity definition. Retrieved from http://medical-dictionary.thefreedictionary.com/obesity, 2003.

[2] World Health Organization. Obesity and Overweight. Retrieved from http://www.who.int/mediacentre/factsheets/fs311/en/, 2016.

[3] National Institute of Diabetes, Digestive, and Kidney Diseases. Overweight and Obesity statistics . Retrieved from https://www.niddk.nih.gov/health-information/health-statistics/Pages/overweight-obesity-statistics.aspx, 2012.

[4] American Veterinary Medical Association. U.S. Pet Ownership statistics. Retrieved from https://www.avma.org/KB/Resources/Statistics/Pages/Market-research-statistics-US-pet-ownership.aspx, 2012.

[5] PD McGreevy, PC Thomson, C Pride, A Fawcett, T Grassi, and B Jones. Prevalence of obesity in dogs examined by australian veterinary practices and the risk factors involved. *Veterinary Record-English Edition*, 156(22):695–701, 2005.

[6] Dictionary. Technology definition. Retrieved from http://www.dictionary.com/browse/technology, 2017.

[7] Vigyanix. Wearable technology: An extension to human body. Retrieved from https://vigyanix.com/blog/wearable-technology-an-extension-to-human-body/, April 2017.

[8] Kiana Tehrani and Andrew Michael. Wearable technology and wearable devices: Everything you need to know. *Wearable Devices Magazine*, 2014.

[9] Ivy Wigmore. Internet of things(iot). Retrieved from http://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT, January 2017.

[10] Animaltalk. Your guide to responsible pet ownership. Retrieved from https://animaltalk.co.za/wp-content/uploads/2015/10/Obesity-Hills-Body-Condition-Score-Chart.jpg, January 2016.

[11] Banfield Pet Hospital. Difference between cats and dogs. Retrieved from https://www.banfield.com/pet-healthcare/additional-resources/article-library/behavior/differences-between-cats-and-dogs, January 2017.

[12] Crunchbase. Findster overview. Retrieved from http://www.dictionary.com/browse/technology, 2017.

[13] Merriam-Webster. Definition of calorie. Retrieved from https://www.merriam-webster.com/dictionary/calorie.

[14] verywell. What is Energy Expenditure. Retrieved from https://www.verywell.com/what-is-energy-expenditure-3496103, 2016.

[15] Elizabeth O'Toole. *Evaluation of the Use of Indirect Calorimetry for the Measurement of Resting Energy Expenditure in Dogs*. University of Guelph, 2000.

[16] Dawn Brooks, Julie Churchill, Karyn Fein, Deborah Linder, Kathryn E Michel, Ken Tudor, Ernie Ward, and Angela Witzel. 2014 aaha weight management guidelines for dogs and cats*. *Journal of the American Animal Hospital Association*, 50(1):1–11, 2014.

[17] David Frankenfield, Lori Roth-Yousey, Charlene Compher, Evidence Analysis Working Group, et al. Comparison of predictive equations for resting metabolic rate in healthy nonobese and obese adults: a systematic review. *Journal of the American Dietetic Association*, 105(5):775–789, 2005.

[18] Emma N Bermingham, David G Thomas, Nicholas J Cave, Penelope J Morris, Richard F Butterwick, and Alexander J German. Energy requirements of adult dogs: a meta-analysis. *PloS one*, 9(10):e109681, 2014.

[19] Compendium of Physical Activities. Corrected mets. Retrieved from https://sites.google.com/site/compendiumofphysicalactivities/corrected-mets, April 2017.

[20] James A Levine. Measurement of energy expenditure. *Public health nutrition*, 8(7a):1123–1132, 2005.

[21] Peter D. Blauner David M. Nunamaker. Methods of gait analysis. Retrieved from http://cal.vet.upenn.edu/projects/saortho/chapter_91/91mast.htm, 2016.

[22] Thomas F. Fletcher Vicki L. Datt. Gait foot-fall patterns. Retrieved from http://vanat.cvm.umn.edu/gaits/, 2012.

[23] Greenhounds. Greyhound breed profile. Retrieved from http://www.greenhounds.com.au/about-greyhounds/greyhound-breed-profile.html, January 2013.

[24] Dr. Ernest Ward. Weight reduction in dogs - general information, 2007.

[25] Mickaël Weber, Thomas Bissot, Eric Servet, Renaud Sergheraert, Vincent Biourge, and Alexander J German. A high-protein, high-fiber diet designed for weight loss improves satiety in dogs. *Journal of veterinary internal medicine*, 21(6):1203–1208, 2007.

[26] Nicholas M Whitney, Harold L Pratt Jr, Theo C Pratt, and Jeffrey C Carrier. Identifying shark mating behaviour using three-dimensional acceleration loggers. *Endangered Species Research*, 10:71–82, 2010.

[27] Ryo Kawabe, Katsuaki Nashimoto, Tomonori Hiraishi, Yasuhiko Naito, and Katsufumi Sato. A new device for monitoring the activity of freely swimming flatfish, japanese flounder paralichthys olivaceus. *Fisheries Science*, 69(1):3–10, 2003.

[28] Paula Martiskainen, Mikko Järvinen, Jukka-Pekka Skön, Jarkko Tiirikainen, Mikko Kolehmainen, and Jaakko Mononen. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Applied animal behaviour science*, 119(1):32–38, 2009.

[29] Maëg Moreau, Stefan Siebert, Andreas Buerkert, and Eva Schlecht. Use of a tri-axial accelerometer for automated recording and classification of goats' grazing behaviour. *Applied Animal Behaviour Science*, 119(3):158–170, 2009.

[30] Junichi Okuyama, Yuuki Kawabata, Yasuhiko Naito, Nobuaki Arai, and Masato Kobayashi. Monitoring beak movements with an acceleration datalogger: a useful technique for assessing the feeding and breathing behaviors of sea turtles. 2010.

[31] Joseph Soltis, Rory P Wilson, Iain Douglas-Hamilton, Fritz Vollrath, Lucy E King, and Anne Savage. Accelerometers in collars identify behavioral states in captive african elephants loxodonta africana. *Endangered Species Research*, 18(3):255–263, 2012.

[32] Dean M Karantonis, Michael R Narayanan, Merryn Mathie, Nigel H Lovell, and Branko G Celler. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE transactions on information technology in biomedicine*, 10(1):156–167, 2006.

[33] Mi Zhang and Alexander A Sawchuk. Motion primitive-based human activity recognition using a bag-of-features approach. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, pages 631–640. ACM, 2012.

[34] Juha Parkka, Miikka Ermes, Panu Korpipaa, Jani Mantyjarvi, Johannes Peltola, and Ilkka Korhonen. Activity classification using realistic data from wearable sensors. *IEEE Transactions on information technology in biomedicine*, 10(1):119–128, 2006.

[35] Fang-Chen Chuang, Ya-Ting C Yang, and Jeen-Shing Wang. Accelerometer-based energy expenditure estimation methods and performance comparison. In *2nd International Conference on Advances in Computer Science and Engineering (CSE 2013). Atlantis Press*, pages 99–103, 2013.

[36] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.

[37] Jonathan Lester, Tanzeem Choudhury, and Gaetano Borriello. A practical approach to recognizing physical activities. In *International Conference on Pervasive Computing*, pages 1–16. Springer, 2006.

[38] Luciana C Jatoba, Ulrich Grossmann, Chistophe Kunze, Jorg Ottenbacher, and Wilhelm Stork. Context-aware mobile health monitoring: Evaluation of different pattern recognition methods for classification of physical activity. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5250–5253. IEEE, 2008.

[39] Annapurna Sharma, Amit Purwar, Young-Dong Lee, Young-Sook Lee, and Wan-Young Chung. Frequency based classification of activities using accelerometer data. In *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on*, pages 150–153. IEEE, 2008.

[40] Chao Chen, Barnan Das, and Diane J Cook. A data mining framework for activity recognition in smart environments. In *Intelligent Environments (IE), 2010 Sixth International Conference on*, pages 80–83. IEEE, 2010.

[41] Whistle. Classifyng Dogs' Activities. Retrieved from http://www.whistle.com/blog/classifying-dogs-activities/, 2013.

[42] Kevin Townsend, Carles Cufí, Robert Davidson, et al. *Getting started with Bluetooth low energy: tools and techniques for low-power networking.* " O'Reilly Media, Inc.", 2014.

[43] Whistle. What is activity tracking? Retrieved from http://support.whistle.com/hc/en-us/articles/209556337-What-is-activity-tracking-, 2013.

[44] Whistle. Can I set an activity goal for my pet? Retrieved from http://support.whistle.com/hc/en-us/articles/209565177-Can-I-set-an-activity-goal-for-my-pet-, 2013.

[45] Whistle. Getting started. Retrieved from http://support.whistle.com/hc/en-us/articles/208674888-How-do-I-set-up-my-Whistle-GPS-Pet-Tracker-on-my-Android-device-, April 2017.

[46] Whistle. How do I reclassify an event? Retrieved from http://support.whistle.com/hc/en-us/articles/209564927-How-do-I-reclassify-an-event-, 2013.

[47] Fast Company. Your dog is fat and lazy: 3 activity trackers for the quantified pet big data is coming to pets. Retrieved from https://www.fastcompany.com/3037268/pet-week/your-dog-is-fat-and-lazy-3-activity-trackers-for-the-quantified-pet, 2014.

[48] MIT management sloan school. A health and fitness tracker for dogs. Retrieved from http://mitsloan.mit.edu/newsroom/articles/a-health-and-fitness-tracker-for-dogs/, 2014.

[49] FitBark. How much exercise does my dog really need? Retrieved from https://www.fitbark.com/insights/exercise/, 2014.

[50] Fitbark. How do i change my dog's activity goal? Retrieved from https://blog.cammy.com/top-smart-pet-collars, December 2016.

[51] Tractive. Features. Retrieved from https://tractive.com/en/gps-2/, 2016.

[52] Cammy team. Top smart pet collars 2017. Retrieved from https://blog.cammy.com/top-smart-pet-collars, January 2017.

[53] Sascha Härtel, Jens-Peter Gnam, Simone Löffler, and Klaus Bös. Estimation of energy expenditure using accelerometers and activity-based energy models—validation of a new device. *European Review of Aging and Physical Activity*, 8(2):109, 2010.

[54] David J Wrigglesworth, Emily S Mort, Sarah L Upton, and Andrew T Miller. Accuracy of the use of triaxial accelerometry for measuring daily activity as a predictor of daily maintenance energy requirement in healthy adult labrador retrievers. *American journal of veterinary research*, 72(9):1151–1155, 2011.

[55] Bernard D Hansen, B Duncan X Lascelles, Bruce W Keene, Allison K Adams, and Andrea E Thomson. Evaluation of an accelerometer for at-home monitoring of spontaneous activity in dogs. *American journal of veterinary research*, 68(5):468–475, 2007.

[56] Kathryn E Michel and Dorothy Cimino Brown. Determination and application of cut points for accelerometer-based activity counts of activities with differing intensity in pet dogs. *American Journal of Veterinary Research*, 72(7):866–870, 2011.

[57] Emmanuel Munguia Tapia. *Using machine learning for real-time activity recognition and estimation of energy expenditure*. PhD thesis, Massachusetts Institute of Technology, 2008.

[58] John G Proakis. Digital signal processing: principles, algorithms, and application-3/e. 1996.

[59] Majid Dadafshar. Accelerometer and gyroscopes sensors: Operation, sensing, and applications. Retrieved from `https://www.maximintegrated.com/en/app-notes/index.mvp/id/5830`, March 2017.

[60] Ryan Morrison. *Physical activity and sedentary behaviour in humans and pet dogs*. PhD thesis, University of Glasgow, 2015.

[61] JB de V Weir. New methods for calculating metabolic rate with special reference to protein metabolism. *The Journal of physiology*, 109(1-2):1, 1949.

[62] HB Williams, JA Riche, and Graham Lusk. Animal calorimetry second paper. metabolism of the dog following the ingestion of meat in large quantity. *Journal of Biological Chemistry*, 12(3):349–376, 1912.

[63] N Gerth, C Ruoß, B Dobenecker, Sven Reese, and JM Starck. Using heart rate to predict energy expenditure in large domestic dogs. *Journal of animal physiology and animal nutrition*, 2015.

[64] Jonathan A Green. The heart rate method for estimating metabolic rate: review and recommendations. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 158(3):287–304, 2011.

[65] Patrick John Butler, Jonathan Andrew Green, IL Boyd, and JR Speakman. Measuring metabolic rate in the field: the pros and cons of the doubly labelled water and heart rate methods. *Functional Ecology*, 18(2):168–183, 2004.

[66] Ann Essner, Rita Sjöström, Erik Ahlgren, and Birgitta Lindmark. Validity and reliability of polar® rs800cx heart rate monitor, measuring heart rate in dogs during standing position and at trot on a treadmill. *Physiology & behavior*, 114:1–5, 2013.

[67] Jiapu Pan and Willis J Tompkins. A real-time qrs detection algorithm. *IEEE transactions on biomedical engineering*, (3):230–236, 1985.

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

[69] L Brouha, WB Cannon, and DB Dill. The heart rate of the sympathectomized dog in rest and exercise. *The Journal of physiology*, 87(4):345, 1936.

[70] Giorgio Roffo and Simone Melzi. Ranking to learn: Feature ranking and selection via eigenvector centrality. *New Frontiers in Mining Complex Patterns, Fifth International workshop, nfMCP2016*, 2017.

[71] G. Roffo, S. Melzi, and M. Cristani. Infinite feature selection. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 4202–4210, Dec 2015.

[72] PPC Chamas, VMC Oliveira, FL Yamaki, and MHMA Larsson. Time-domain signal-averaged electrocardiogram in healthy german shepherd and boxer dogs. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, 66(3):778–786, 2014.

[73] Alexander Campbell and Michael Chapman. *Handbook of poisoning in dogs and cats*. John Wiley & Sons, 2008.

[74] Robert L Van Citters and Dean L Franklin. Cardiovascular performance of alaska sled dogs during exercise. *Circulation Research*, 24(1):33–42, 1969.

[75] Ivalyo Christov and G Bortolan. Ranking of pattern recognition parameters for premature ventricular contractions classification by neural networks. *Physiological Measurement*, 25(5):1281, 2004.

[76] Allison P Lamb, Kathryn M Meurs, and Robert L Hamlin. Correlation of heart rate to body weight in apparently normal dogs. *Journal of Veterinary Cardiology*, 12(2):107–110, 2010.

[77] MJ Hezzell, K Humm, SG Dennis, L Agee, and A Boswood. Relationships between heart rate and age, bodyweight and breed in 10,849 dogs. *Journal of Small Animal Practice*, 54(6):318–324, 2013.

[78] Jonathan M Yashari, Colleen G Duncan, and Felix M Duerr. Evaluation of a novel canine activity monitor for at-home physical activity analysis. *BMC veterinary research*, 11(1):146, 2015.

Appendix A

User's Manual

User's Manual
Manual de uso e instalação da aplicação de teste

João Sousa

July 26, 2017

1 Instalação do ficheiro apk no telemóvel

Windows

- Em windows, simplesmente deve ligar o seu smartphone (com sistema operativo Android) ao computador através de um cabo USB, ir a Meu Computador e aceder ao dispositivo
- De seguida deve aceder ao link *https://www.dropbox.com/s/pcovi7ht6ppmpeb/Application-debug-unaligned.apk?dl=0* e premir o botão download para descarregar o ficheiro apk
- Finalmente deve copiar ou arrastar o ficheiro apk que descarregou e colocá-lo numa diretoria qualquer do smartphone (aconselhavelmente uma diretoria que seja facilmente identificada por si)

Mac

- Primeiro deve aceder ao site *https://www.android.com/filetransfer/* e premir o botão "Download Now".
  Para instalar o Android File Transfer basta arrastar o ícone para a pasta aplicações.
- De seguida, deve aceder ao link *https://www.dropbox.com/s/vvbg3zjunynjg7k/Application-debug-unaligned.apk?dl=0* e premir o botão download para descarregar o ficheiro apk
- Uma vez obtido o ficheiro apk, deve ligar o seu smartphone (com sistema operativo Android) ao computador através de um cabo USB, e só quando estiver ligado, abrir a aplicação Android File Transfer.
  Possivelmente a aplicação poderá abrir automaticamente e gerar uma imagem semelhante à seguinte.

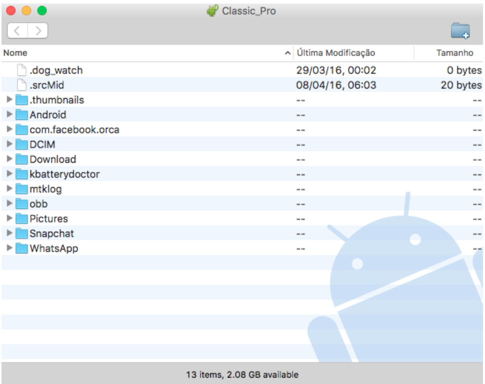
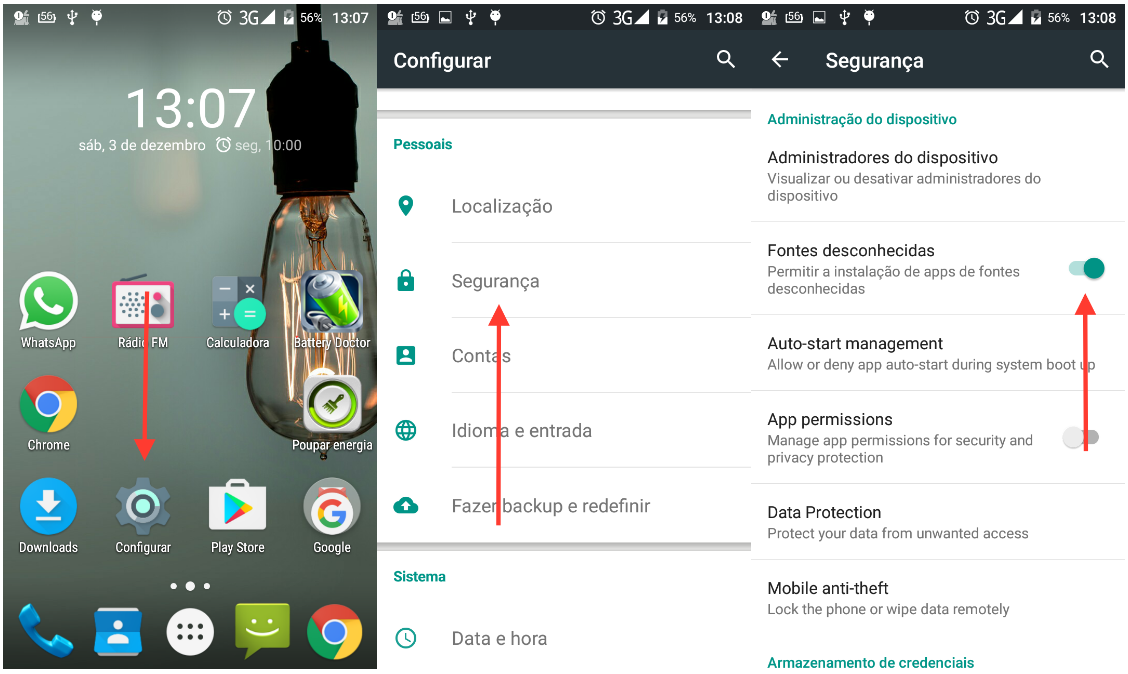


Figure 1: Janela gerada pelo Android File Transfer

- Finalmente deve copiar ou arrastar o ficheiro apk que descarregou e colocá-lo numa diretoria qualquer do smartphone (aconselhavelmente uma diretoria que seja facilmente identificada por si)

2 Autorizar instalação de aplicações externas

- Uma vez no seu smartphone deve dirigir-se a Configurar ≫ Segurança ≫ Fontes desconhecidas e ligar o botão que permite a instalação de apps de fontes desconhecidas.



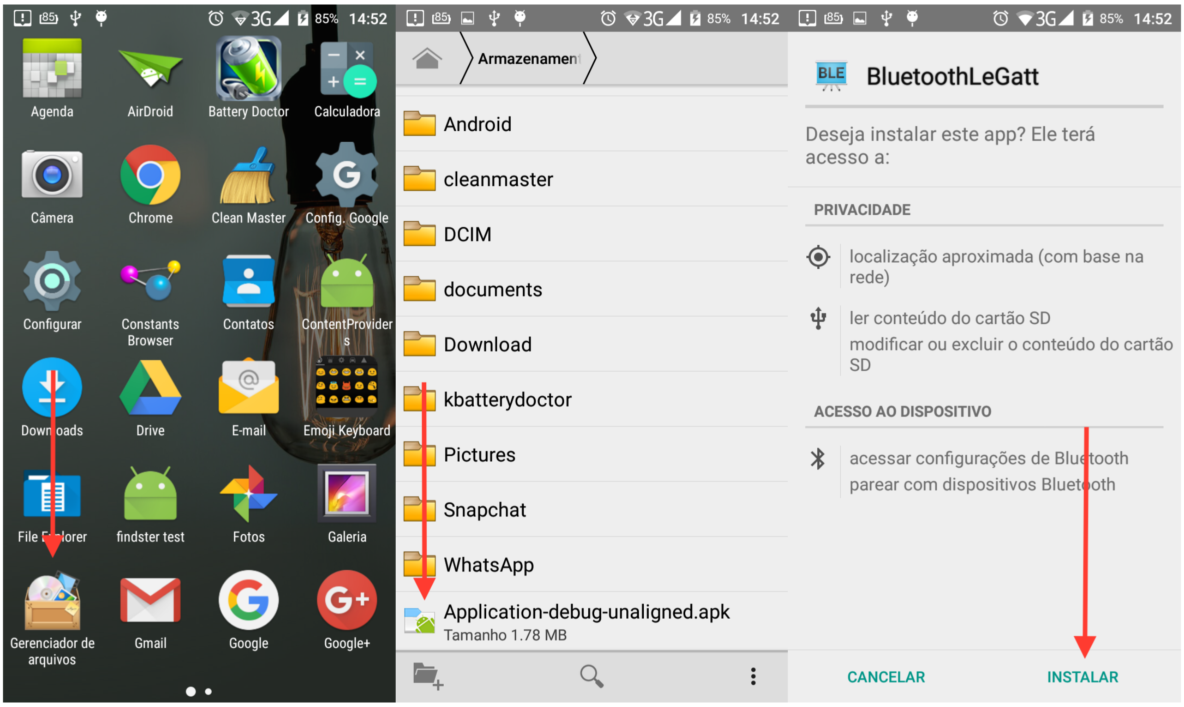
(a) (b) (c)

Figure 2: Passos a seguir

3 Instalação da aplicação

- Para aceder à aplicação copiada deve abrir a aplicação que permite gerir as pastas e ficheiros do seu smartphone, tipicamente chamada File Manager ou Gerenciador de Arquivos

- A partir dessa aplicação deve encontrar o ficheiro apk que copiou anteriormente e clicar sobre ele. Ser-lhe-á perguntado se deseja instalar o aplicativo. Clique que sim e terá a app instalada no seu smartphone.



(a) (b) (c)

Figure 3: Instalação do apk

4 Utilização do sistema e aplicacão

Sistema

- Antes de poder começar a utilizar a aplicação assegure-se que os módulos foram carregados durante algum tempo (aproximadamente 30 minutos / 1 hora).
  Para o fazer ligue o cabo USB ao carregador (peça branca) e ao computador.

- De seguida deverá pegar no módulo azul (Rx) e carregar no seu botão até que uma luz acenda.
  Quando o mesmo acontecer, o módulo estará pronto para funcionar.

- Finalmente deve carregar no botão do módulo amarelo (Tx) até que a sua luz verde desapareça. O objetivo é que o módulo azul esteja a piscar intermitentemente! Caso isso não aconteça carregue mais uma vez no módulo azul.

5 BluetoothLEGatt

- Numa primeira fase, o utilizador deve ligar o bluetooth no seu smartphone a partir de Configurar ≫ Bluetooh e mudar o desativado para ativo.

- Agora sim aceda à aplicação que deve gerar uma imagem semelhante a esta.

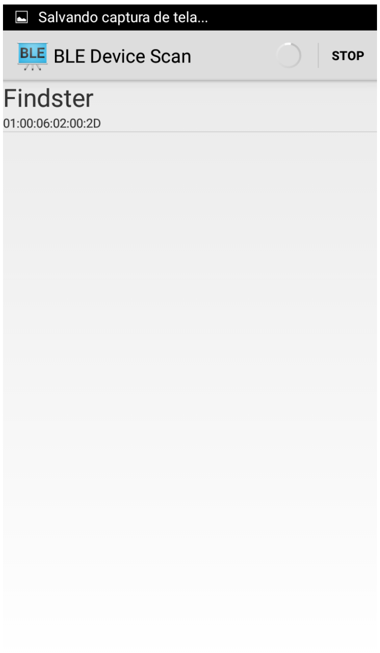
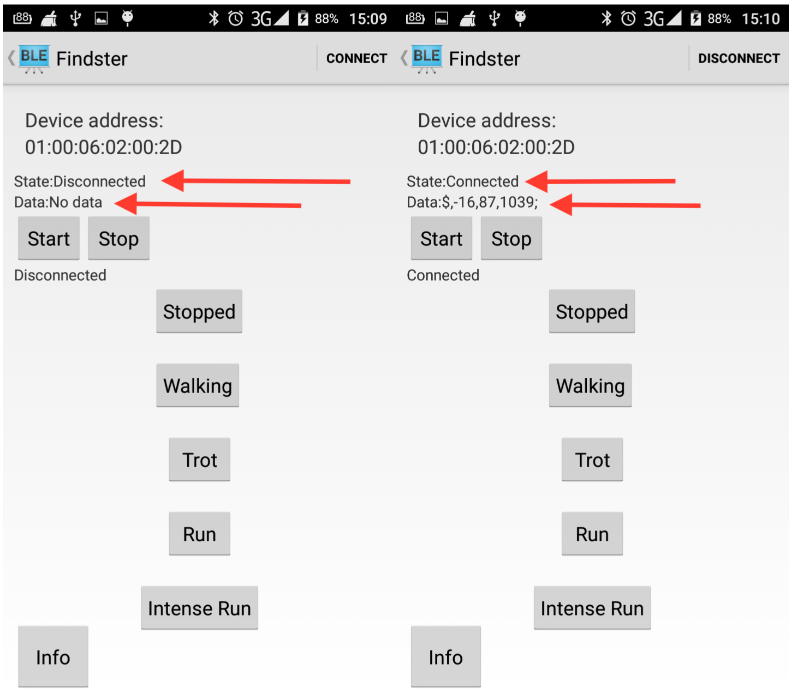


Figure 4: Janela gerada pela aplicação

Se tudo estiver funcional, o aparelho Findster deve aparecer e deve carregar no mesmo. Caso o Findster não apareça é porque o módulo azul não está iniciado corretamente, pelo que deve voltar os passos anteriores do tópico 4.

- Uma vez selecionado o dispositivo uma das seguintes imagens deve aparecer.



(a) (b)

Figure 5: Menu da aplicação

Automaticamente os aparelhos irão ligar-se ao smartphone, por isso apenas deve esperar até que a linha "Data" apresente valores que estão constantemente a mudar, ou que a linha "State" indique "Connected"

- Finalmente deve carregar no botão "Start" para começar a gravar, "Stop" para interromper a gravação (pode servir como pausa).
  Os outros 5 botões devem ser premidos quando o utilizador perceber que o cão está num dos referidos estados, por exemplo, se o mesmo estiver parado, deve premir-se o botão "Parado", caso esteja num andar lento, "Andar", num andar acelerado, "Trote", etc.

- Por último mas não menos importante, deve colocar o dispositivo amarelo (e não o azul!) na coleira do seu cão e estará pronto para as medições!

Caso exista algum problema, não hesite em contactar:
e-mail: jlbgsousa@gmail.com
facebook: https://www.facebook.com/joao.garcia.73550794

Appendix B

Equipment Setup

Equipment Setup
Manual de instruções

Televet Holter monitoring

1 Configuração do Findster

Coloque o módulo Findster na coleira de forma firme.

Coloque a coleira no animal de forma firme mas confortável para o animal.

2 Configuração do Holter

2.1 Equipamento

Verifique se todo o material necessário está reunido, nomeadamente:

- Monitor
- 4 Elétrodos de ECG (mínimo)
- Gel condutivo de ECG
- Cartão de memória formatado
- Pilhas alcalinas
- Ligaduras
- Tesoura
- Fita adesiva
- Folha indicativa do local de aplicação de cada elétrodo
- Smartphone com 30% de bateria (mínimo)

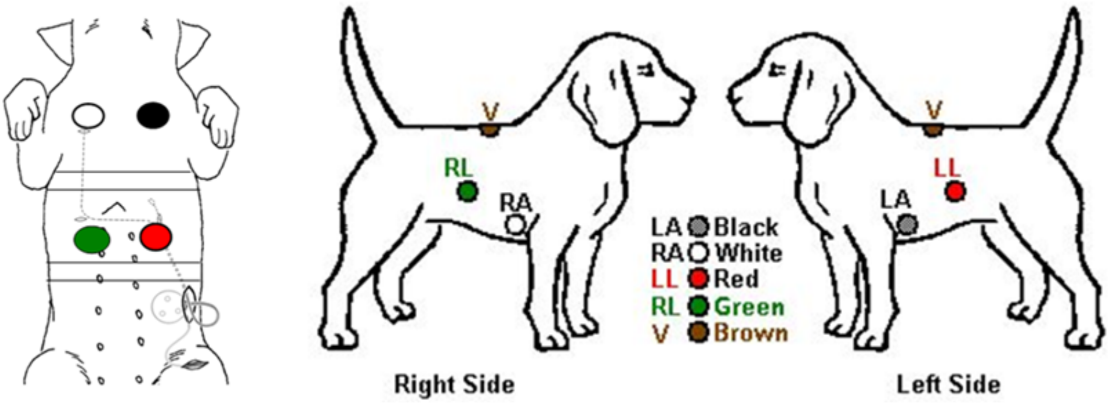
- Smartphone com aplicação BluetoothLeGatt
- Módulo Findster Child
- Módulo Findster Guardian
- Espaço alargado para o animal caminhar

2.2 Procedimento

Coloque as pilhas alcalinas no monitor assim como o cartão de memória formatado. De seguida coloque o monitor na sua bolsa própria.

Corte 5 tiras de fita adesiva com a tesoura.

Humidifique a zona de aplicação dos elétrodos com gel de condutividade apropriado e coloque os quatro elétrodos no animal de acordo com a folha fornecida no kit do monitor holter. Se possível evite que o animal se deite.



Ligue os elétrodos aos fios condutores e, se necessário, para reduzir a sua liberdade prenda-os com as tiras de fita adesiva.

Antes de ligar os fios ao monitor holter, envolva os fios com uma ligadura ou outro material de forma a pressionar os elétrodos e envolva bem o animal.

Coloque o monitor no dorso do animal por cima da ligadura e fixe-o da melhor forma possível.

Ligue os fios ao monitor e prenda-os com as tiras de fita adesiva de forma a reduzir o seu movimento. Uma vez ligados os fios ao monitor uma luz vermelha do mesmo deve acender indicando que a medição do holter for iniciada.

No smartphone active o bluetooth e abra a aplicação BluetoothLeGatt. Faça o emparelhamento com o módulo Guardian, dê um nome ao teste e pressione play.

3 Fase de Teste

Uma vez que todo o sistema esteja preparado, o animal deve permanecer em repouso durante pelo menos 5 minutos (sentado ou deitado é indiferente).

De seguida o animal deve caminhar a um ritmo constante durante pelo menos 5 minutos.

Passado esse tempo, o ritmo de caminhada deve aumentar de forma a que o animal esteja em ritmo de trote. Neste caso não é crucial que a velocidade seja constante, mas sim que o animal apresente esforço durante a atividade. Um período mínimo de 5 minutos é necessário.

Após a fase de trote o animal deve retornar a um ritmo mais lento de caminhada durante 5 ou mais minutos e finalmente deve repousar, agora durante um período mais alargado de 10 minutos no mínimo.

Finalmente termine a medição ao mesmo tempo nos dois dispositvos carregando no botão stop da aplicação ao mesmo tempo que retira os fios do monitor holter

Appendix C

Test Protocol

Protocolo de teste

Informação do animal	
Nome:	
Raça:	
Porte:	
Idade:	
Peso (Kg):	

Teste
Dia (aa/mm/dd):
Hora de início:
Hora de fim:

Fase	Ação	Resultado (S,N)	Comentário
1	Período de repouso (pelo menos 5 minutos)		
2	Período de caminhada, velocidade lenta (5 minutos)		
3	Período de trote (pelo menos 2 minutos)		
4	Período de corrida (alternar com trote, 2 minutos)		
5	Repetir a fase número 3		
6	Repetir a fase número 2		
7	Período de repouso (pelo menos 10 minutos)		