School of Natural &
Environmental Sciences

The use of wearable sensors for animal behaviour assessment

Robin J. Thompson

A thesis submitted for the degree of Doctor of Philosophy (PhD)

February 2018

Abstract

The research outlined in this thesis presents novel applications of wearable sensors in the domain of animal behaviour assessment. The use of wearable sensing technology, and in particular accelerometry, has become a mainstay of behaviour assessment in humans, allowing for detailed analysis of movement based behaviour and health monitoring. In this thesis we look to apply these methodologies to animals and identify approaches towards monitoring their health and wellbeing. We investigate the use of the technology in the animal domain through a series of studies examining the problem across multiple species and in increasingly complex scenarios. A tightly constrained scenario is presented initially, in which horse behaviour was classified and assessed in the context of dressage performances. The assessment of lying behaviour in periparturient sows confined to gestation crates examines a scenario in which the movement of the subject was constrained, but not predetermined. Expanding this work to include sows housed in free-farrowing environments removed the movement constraints imposed by the gestation crates. We examine the implications of the use of multiple sensors and how this might affect the accuracy of the assessments. Finally, a system for behaviour recognition and assessment was developed for domestic cats. Study animals were free to move and behave at their own discretion whilst being monitored through the use of wearable sensors, in the least constrained of the studies. The scenarios outlined herein describe applications with an increasing level of complexity through the removal of constraints. Through this work we demonstrate that these techniques are applicable across species and hold value for the wellbeing of both commercial and companion animals.

Declaration

This thesis has been composed by myself and has not been submitted as part of any previous application for a degree. All sources of information have been specifically acknowledged by means of referencing.

Robin Thompson

Publications

- Thompson, R., Kyriazakis, I., Holden, A., Olivier, P. and Plötz, T. (2015) 'Dancing with horses', in Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15, pp. 325-336. doi: 10.1145/2750858.2807536.

- Thompson, R., Matheson, S. M., Plötz, T., Edwards, S. A. and Kyriazakis, I. (2016) 'Porcine lie detectors: Automatic quantification of posture state and transitions in sows using inertial sensors', Computers and Electronics in Agriculture, 127, pp. 521-530. doi: 10.1016/j.compag.2016.07.017.

Acknowledgements

I would like to extend my most sincere thanks to my supervisory team, Ilias Kyriazakis, Thomas Plötz, and Stephen Matthews, for their support and guidance over the past four years. Whilst at times things were difficult, knowing that you were always looking over my shoulder gave me the necessary boost to keep going.

I would not have even been able to consider starting a PhD, let alone reaching the end of one, if it were not for the love and support of my family. When things were hardest you all gave me the strength to push through.

To Hugo, Lisa, and Wojciech, I thank you wholeheartedly for putting up with me. I don't deserve such good friends, and I am thankful I have you all.

And finally thanks go to my wonderful proofreaders, particularly Lydia, who insists she actually learned something.

This thesis has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number 613574 (PROHEALTH). This project has also received funding from the Biotechnology and Biological Sciences Research Council (BBSRC) in the form of a studentship.

Contents

List of Figures

List of Tables

1

Introduction

Contents

In this thesis, we develop techniques and methodologies for the automatic assessment of animal behaviour. The development and application of sensors has enabled the automated monitoring of animals both at large scale and over substantial periods of time. We present a series of case-studies in which we describe the challenges associated with automating this process. Through these studies, we highlight methods by which we can improve the productivity, health and welfare status, and understanding of both livestock and companion animals.

Automated monitoring and assessment of animal behaviour provides avenues by which the requirement for human interaction can be reduced, whilst maintaining high standards of animal care. This, in turn, allows for a more sustainable intensification of farming practices, as cost and time investment are reduced, whilst high standards of animal welfare are maintained. In the case of livestock, this is consistent with the trend for production systems that involve large numbers of animals, whilst reducing the contact between the animal and its keeper [49]. The techniques we have developed are not, however, restricted to the observation of livestock. Animals often perform important roles in their interactions with humans. For example, guide dogs are able to lead vision impaired people, and horses provide a means of travel or entertainment [45]. Understanding how these animals perform these tasks and the quality with which they do so, can lead us to improve the experience of their human companion. Not all applications necessarily seek to exclusively benefit the health of animals. There is a current trend towards “life-logging” [128], a practice in which people monitor a wide range of their daily activities, with a view to improve aspects of, for example, their health and productivity based on the information gained. This trend even extends to observing the activities and behaviours of our pets, with the aim of improving their wellbeing, and understanding how they act in our absence.

This introductory chapter aims to provide insight into the benefits of automatic monitoring of animal behaviour and raise issues associated with its implementation. We shed light on applications of behaviour monitoring and how automation can reduce the resources required, whilst reducing subjectivity and improving consistency. We explore different approaches to automation and examine the benefits of wearable sensors over other technologies. Finally, we provide an overview of the structure of the thesis and define the narrative that brings experimentation with three different species together.

1.1 Animal Welfare

We rely on livestock to provide us with food, fuel and clothing, and are constantly seeking techniques to increase their productivity. There is, however, a balance to be found between this increased productivity and the sacrifices to animal health and welfare that so often accompany it. The concept of animal welfare can be described in several different ways, however it has been refined and formalised by Stamp-Dawkins [39] in which the author suggests that the welfare of an animal can be assessed through the answers to just two questions: “are the animals healthy?”, and “do they have what they want?”. Whilst these are straightforward questions, the answers are not always easy to ascertain and they reveal the core concepts that need to be addressed when considering how best to manage the animals under our care.

The question, “Are the animals healthy?”, is the cornerstone of veterinary practice. The identification of challenges to an animal’s health is often straightforward. Through the observation and assessment of clinical symptoms of a condition, experienced individuals are able to provide a diagnosis and recommend appropriate treatment. Many conditions initially present sub-clinically, that is, without clinical symptoms. It has been established that behavioural changes may precede the onset of clinical symptoms [60, 85, 150], the observation of which may aid in diagnosis [50, 81]. Such changes include subtle changes in their feeding and drinking behaviour, or their locomotory behaviour. In many cases the diagnosis of a health challenge prior to the exhibition of clinical symptoms can lead to improved prognosis for the animal, improving the welfare of the animal and reducing loss of productivity [133].

The second question proposed by Dawkins, “Do the animals have what they want?” is a more difficult question to answer. Besides the necessity to determine what we deem the animal to want, the ability of animals to communicate their satisfaction to those responsible for their care is limited. Fraser [47] suggests that an animal should have “positive experiences, such as comfort and contentment”. How animals experience emotion cannot be measured directly [104], however behavioural approaches have been suggested [123]. If we are to continue to increase the productivity of commercial animals, it is essential that we consider their welfare, both physical and emotional, as a priority. Consequently, it is apparent that a clear understanding of animal behaviour can help us to reach this target.

1.1.1 High value individuals - not all animals equal

In the context of commercialisation, we quantify the value of an animal in terms of monetary cost. Consequently, some animals hold more value than others. Animals we are able to breed quickly and in large numbers, and are housed for short periods before slaughter, such as chickens or pigs bred for meat, hold less value than those whose raising requires significant time and resources, such as cows and breeding sows. Subsequently there is a decision to be made regarding the investment we make into these animals. The approaches we have developed in this thesis specifically target these high value animals, as those animals with lower individual value may not warrant the investment. We similarly consider companion animals to be of high value, but for different reasons.

1.2 Monitoring Animal Behaviour

The monitoring of animal behaviour is a well-established field of research and many approaches have been developed and implemented, both in terms of manual and automated observation. Traditionally, observation of behaviour is conducted by an invested individual whose task it is to record the occurrences and durations of salient behaviours exhibited by the subject [131]. This, however, is prone to subjectivity and inter-observer variability, as well as being time consuming and expensive [61]. Even so, much research has been performed to identify behavioural patterns that can inform us about the underlying states of the animals. The study of sickness behaviour, for example, aims to identify behavioural patterns indicative of a challenge to the animals' health. Hart [59] explains that, at the time of writing, research into the typical behavioural patterns of animals focussed on that of healthy individuals, and consequently overlooks what can be learned from examining the manner in which an animal responds to sickness. Whilst this was true at the time, more recent literature has sought to quantify this behaviour, and use it as an avenue for identifying illness in animals. It has been found that a range of behavioural patterns can be interpreted to provide insight into the health of an animal. Quantification of pain, for example, has clear welfare outcomes. As we are unable to communicate directly with the animal, we must rely on behavioural changes to identify pain. Keegan [71] for example, identified behavioural responses to pain in horses suffering from lameness, suggesting that, depending on the location of the pain, the horse will respond in different ways. Similarly, [127] validated approaches to assessing the gait of cattle, and noted improvement of gait scores on the administration of analgesics. Research has also identified behavioural changes associated with an infection. [41] describe changes in social exploration associated with infection in laboratory rats, and [66] identify changes in activity levels in pigs. Whilst these challenges to the animals' wellbeing are physiological, there is also a need to address challenges to the animals' mental health. A common expression of mental distress in animals is the exhibition of stereotypic behaviours, that is, the "repetitive and invariant" display of behaviours "that have no obvious goal or function" [99].

In addition to behaviour changes caused by ill health, it is possible to use behavioural changes as indicators of other productivity-affecting factors. Feather pecking, for example, describes a problem faced in commercial chicken units in which chickens peck at each other's feathers and skin. [89] described a method for predicting outbreaks of feather pecking using computer vision techniques, allowing identification of factors promoting this behaviour. Tail-biting behaviour is similarly damaging in pig units [42]. Approaches to predicting this behaviour have focussed on identifying group behaviours that may be indicators that tail biting will occur [132, 139], but have produced mixed results.

The above examples highlight the benefits associated with using behaviour as predictors of welfare challenges and as a diagnostic tool following the onset of a challenge. The majority of these examples rely on some form of subjective assessment, generally delivered by an expert in the field. It is possible that through the automation of behaviour monitoring and assessment more objective metrics could be established, based on quantitative data extracted from observations. Even given the use of an initial subjective assessment to establish ground truth, automation allows for consistency in all following evaluation.

1.3 Automated Techniques for Monitoring Animal Behaviour

With advances in technology brought about through the ubiquity of comparatively inexpensive sensing platforms, strides have been made in recent years towards automating the process of monitoring animal behaviour. A range of technologies are particularly suited to the observation of behaviour [102]. Perhaps the most intuitive of these is computer vision, that is, the techniques involved in automatically interpreting digital images to describe the behaviour exhibited by the subjects.

Computer vision has been used extensively to measure animal behaviour in a range of applications, and has found prominent use in agricultural contexts [38]. Extensive work has been conducted into the use of computer vision to assess equine gait and behaviour [25], often employing markers secured to the animal to aid in tracking [88]. Computer vision for the assessment of porcine behaviour is becoming well established, and has been employed for several applications including monitoring aggressive behaviour [143], activity levels [112] and water consumption [68], for example. Analysis of two-dimensional video data is not always robust to movements in three-dimensional space however. Changes in posture and movement between vertical levels are often difficult to capture using standard computer vision applications. As a result, investigations into the use of depth imaging have been conducted [103, 147]. The above approaches to automatically monitoring animal behaviour generally rely on the animal under observation being in a fixed or predictable location, where monitoring technology has been set up.

Computer vision and depth imaging are particularly well suited to monitoring groups of animals. This has particular value in the assessment of the behaviour of “low value” stock, such as chickens, as the investment per animal is relatively low. The cost of these approaches increases as the number of animals under observation decreases, however. Given a high value individual, such as a breeding sow or dairy cow, it becomes reasonable to consider other options for monitoring their individual behaviour. When only a single subject is to be considered, it becomes practical to invest more initially and allows a more direct approach to sensing. With the reduction in size and cost of sensors a move is being made towards wearable sensing [126], that is, sensors that are secured to the animal, allowing behaviour to be recorded regardless of the location of the animal. Irrespective of the sensing solution employed for monitoring behaviour, a system to automatically determine what behaviours are being displayed is required to automate the process.

1.3.1 Wearable sensors

A common problem of in-place sensing solutions, such as cameras is the tracking of subjects over large ranges. This can be addressed through the use of wearable sensors. The use of radiotelemetry for animal tracking is well established [28] and has been used extensively to monitor population distributions, resource use, and demographic changes of groups of animals in wild environments [129]. The miniaturisation and cost reduction of GPS sensors has further expanded the opportunities available to animal ecology researchers [21]. Whilst these sensors provide details of the gross movements of animals, other technologies are required to perform an assessment of fine locomotor behaviours.

1.3.2 Accelerometers

This thesis focuses on the use of an accelerometer to record the movements of subject animals. An accelerometer is a device capable of measuring acceleration and comes in several forms; however in this work we employ a MEMS (microelectronic systems) piezoelectric (PE) accelerometer [8]. The MEMS accelerometer operates by measuring voltage change caused by changes in stress on a piezoelectric material caused by a pressure from loading mass restricted to movement in a single direction. As the sensor accelerates the loading mass induces a stress on the piezoelectric material causing the voltage to change, see Figure 1.1.

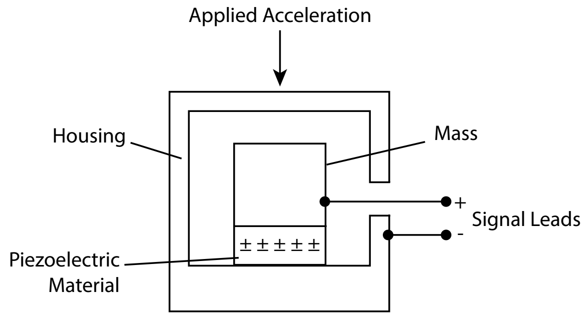


Figure 1.1: Overview of a piezoelectric accelerometer. An applied acceleration causes the loading mass to exert stress on the piezoelectric element, causing a change in voltage which can be read from the signal leads and interpreted to measure the acceleration of the sensor.

The use of a single PE accelerometer restricts measurements to a single spatial axis, however the Axivity AX3 [8] used in the studies in this thesis employs three of these sensors to allow measurement of acceleration in all three spatial dimensions. Accelerometers produce recordings at a predetermined sampling rate and sensitivity. The sensitivity of the accelerometer readings refers to the range of accelerations that the sensor is able to capture. The AX3 sensor operates at a default setting of $\pm 8g_0$ ($\pm$78.48 m/s), allowing it to capture all but the most extreme accelerations exerted on the sensor. The sampling rate of the sensor refers to the number of recordings taken every second. Work has been conducted to establish the optimum sampling rate for measuring human behaviour [75], in which it was demonstrated that optimal sampling rates were application dependant, and that higher than necessary sampling rates could lead to over use of battery power and consequently restrict the potential length of studies.

The nature of the accelerometer makes it particularly well suited for use as a tool for automating the monitoring of movement-based animal behaviour. Due to its small and lightweight form factor, low power usage and relative durability, it is possible to deploy these sensors in contexts that would otherwise be impractical through traditional approaches to behaviour monitoring.

Accelerometers have been used extensively to monitor the behaviour of animals in laboratory environments, with applications such as measuring head movements during brain activity recording [80, 108, 109]. The versatility of body-worn accelerometers, however, makes them ideal for use outside of a laboratory setting. In agricultural contexts, accelerometry has been used to measure aspects of animal behaviour such as gait and lameness detection [53, 101, 116], posture [100, 124, 148] and activity levels

[34, 111, 156]. Analysis of more complex, compound behaviours present a more challenging problem for automated analysis. Aggression and 'play' behaviour, for example, often involve multiple subjects and series of behaviours, requiring a more sophisticated approach to assessment and, potentially, the introduction of a multimodal approach to sensing.

A primary factor affecting the data collected by an accelerometer is the position at which it is secured to the subject. Depending on the location of the sensor, measurements will vary significantly. Research has been conducted into the impact that sensor location has on the assessment of human movement [27]; however, there is little literature concerning this in animal movement assessment. The choice of sensor placement must take into account the movements being observed as well as the anatomy of the subject. In studies on pig movement, common sensor locations include the neck (on a collar) [31], the ear (as part of an ear tag) [111] or the leg [124]. These locations each suffer from their own drawbacks, however each is well justified. Lameness assessment has also been investigated using accelerometry, in horses [73], cattle [116] and pigs [53]. Lameness presents as an abnormality of gait or stance, and as such it would seem intuitive to measure the movement of the legs of the affected animal. Understanding how lameness affects the movement of the animal from other perspectives however may give an indication to the cause of the condition. Sensors secured to a horse's head were able to quantify asymmetry of movements for use in detecting forelimb lameness [71]. Whilst data collected from the animal is certainly informative, varying the sensor location may provide insight into the assessment of lameness.

Considerations regarding the impact on the subject's behaviour due to sensor placement must also be made. As we describe in this thesis, cats unused to wearing a collar exhibited increased scratching behaviours (see Chapter 5) and attempts to secure a sensor to the noseband of dressage horses led to excessive shying away (see Chapter 6). Further investigation into the impact changing the location of the sensor has on the quality of the data, as well any impact it may have on the behaviour of the subject has the capacity to improve study design and outcomes, and constitutes a significant effort in this thesis.

1.3.3 Gyroscopes

In addition to accelerometery, we also employ gyroscopes in Chapter 2. Where accelerometers measure acceleration, gyroscopes measure rotational velocity. This provides us with the ability to measure the rotation of the sensor around the three spatial axes. Depending on the positioning of the sensor this can provide valuable insight into the movement, and consequently the behaviour, of the subject. The gyroscope has certain limitations however, primary amongst which is power usage. At the time of writing, modern MEMS gyroscopes drain device batteries up to twenty times faster than an accelerometer alone. This means that gyroscopy is only practical for collection in short duration studies, such as that described in Chapter 2, however for observational studies designed to run over periods of weeks, it is not feasible to collect this kind of data.

1.3.4 Activity recognition

Activity recognition describes the process of using machine learning techniques to automatically classify data, recorded from an observed subject, according to the activity being displayed. Activity recognition has its roots in the field of Human Computer Interaction (HCI) [20], in which it has been used extensively to provide users with information to assist in the execution of tasks [1]. Bulling et al. [20] describe the process of activity recognition using inertial sensors in detail regarding human activity recognition. In this work a generally applicable sequence of steps to approach the development of activity recognition algorithms is outlined. This sequence, or "pipeline", takes an input of raw inertial data and outputs labels denoting the activity described by the data. Considered in the context of animal activities, approaches have been developed for a range of tasks: in dogs [82] to identify posture and behaviours; cats [156] to create a diary of their behaviours; cattle [148] to distinguish between feeding, chewing and resting behaviours, and pigs [31] to classify posture and behaviour, to name a few. The collection of data from companion animals is gaining relevance as owners and veterinarians strive to gain a better understanding of their pets' activities and health. Research into activity recognition in domestic animals has been conducted in canine subjects [82, 140, 141], and indeed is now being commercialised. The development of activity recognition algorithms generally requires the application of domain specific knowledge concerning the behaviours of the subject. Consequently, developing activity recognition-based applications requires research into identifying appropriate parameters, and establishing expectations.

1.3.5 Movement characteristic assessment

Beyond activity recognition, valuable data relating to the behaviour of the animal can be gained by analysing the characteristics of said behaviours. Where activity recognition is able to provide information regarding the routines and habits of the animals under observation, movement characteristic assessment allows us to differentiate between instances of behaviours and identify aspects of the behaviour to provide us with insight into the underlying state of the subject. In the context of human behaviour this has been applied to the field of skill assessment using accelerometry [76]. However, in animals we have the opportunity to explore parameters that manual observation, and instance recording alone cannot provide. As described above, lameness detection has been performed in horses [71] and cattle [116]. This requires the comparison of movement characteristics associated with walking behaviours in an affected subject with a healthy example. The high sample rate measurements achievable through the use of accelerometers afford us the ability to inspect the movements of the subject in great detail and identify characteristics that may be indicative of a health challenge. Movement characteristic assessment may also give us the ability to differentiate between animals' performances in certain tasks. Research is currently being conducted into assessing the suitability of working dogs [3]. The use of movement characteristic assessment could find use as a tool for identifying animals that respond quickly to owner commands and draw objective comparisons between dogs. Other working animals could similarly be assessed in this way.

1.4 Problem Statement and Thesis Structure

This thesis aims to identify and develop solutions to the technical and practical challenges associated with wearable sensors for monitoring and assessing animal behaviour, in both livestock and companion animals. We examine the problem from a perspective of increasing behavioural complexity seeking to identify areas in which the use of wearable sensors excels, and where they may be found to be unsuitable. We consider two forms of behavioural complexity: the complexity associated with the animals' ability to express natural behaviours and the complexity of the movements involved in the exhibition of said behaviours. When developing approaches to measuring the behaviour of animals a consideration towards real-world applications must be made.

As described above, monitoring of animal behaviour in laboratory conditions is an established field of research. Experiments conducted in these settings represent, through necessity, a tightly constrained scenario, in which the subject is often restricted in terms of movement and ability to express natural behaviours. Whilst these experiments provide a basis for understanding the approaches and techniques required for automated monitoring, challenges associated with implementation outside of the laboratory are not considered. It is the aim of this thesis to build upon the foundational work performed in laboratory studies, whilst systematically relaxing the constraints to identify and propose solutions to the challenges that arise. The four experimental chapters of this thesis document a series of scenarios in which these constraints are progressively removed, both in terms of the movements of the subjects, and their ability to exhibit behaviours in a natural manner.

The first experimental chapter describes the use of wearable sensors to assess the performance of a dressage horse and rider. Dressage requires the horse to perform a predefined set of locomotory behaviours (movements), such as a trotted circle, changes in paces or more complex tasks such as the piaffe; the performance of these movements is then assessed by trained judges [110]. We hypothesise that we are able to quantify an animal's performance and use it as a tool to objectively differentiate between animals exhibiting similar behaviours. The subject in this scenario is the horse, to whom several sensors have been secured. We consider this to be a tightly constrained scenario in which the horse has very little freedom to exhibit natural behaviours as its movements are being informed by the commands of the rider. Further, the series movements that the horse is expected to perform are predetermined, allowing for an assessment of the characteristics of said movements without requiring prior activity recognition in terms of these movements.

Chapters 3 and 4 examine the efficacy of wearable sensors for improving welfare and productivity of sows in a commercial unit. We hypothesise that we are able to identify patterns of animal behaviour that reveal underlying traits. The first of these studies, in which movement assessment is performed on periparturient (the period around giving birth) sows housed in gestation pens, operates on a scenario in which the subjects are constrained in terms of their movements and their ability to express natural behaviours. This restriction is a consequence of the system of production, as opposed to an artificial one. Unlike the previous study, however, in this context the animal's movements are not predetermined; it is only the behavioural repertoire that is restricted. This provides us with a lessening of the constraints and introduces a requirement for a more in-depth approach to activity recognition, one in which we consider the posture of the subjects at all times. We expand upon the activity recognition pipeline by performing an assessment of the movements exhibited by the subject, with the aim of identifying characteristics that relate to the welfare of the animals.

The second investigation into the lying behaviour of sows removes the constraints on movement imposed on the subjects. We consider an implementation for movement assessment in periparturient sows housed in a loose-farrowing environment [22]. In this setting the sow is able to move with a much greater degree of freedom and is able to express some natural behaviours. Thus, the problem is expanded to include an assessment, not only of posture state and transitions, but also active behaviours such as feeding and walking. This represents a further lifting of constraints and, through this, we aim to shed light on the use of multiple sensors as an approach to improving performance in line with the increased complexity of the problem.

The final scenario, described in Chapter 5, looks at a situation where behavioural and movement constraints are removed altogether. We hypothesise that the routines of cats are interpretable through activity monitoring and that, through the identification of salient behaviours, health and welfare assessments may be made. We consider the application of wearable sensors towards the monitoring and assessment of feline behaviour in a domestic setting. The subjects under observation were given full autonomy of behaviour and provided the freedom to move as they pleased. This necessitates a full activity recognition pipeline in which the complex behaviour and activity patterns of the subjects are assessed.

2

Dancing with Horses

Automated Quality Feedback for Dressage Riders

Contents

2.1 Introduction

The equestrian practice of dressage is centuries old. As early as 450 BC, Athenian historian and soldier Xenophon described the "selection, care and training of horses in general" [155]. With roots as a training method for war horses, dressage has progressed and been refined to its current status as an Olympic sport [110]. In the UK alone 17 in $1,000$ people own at least one horse totalling to approximately 1m horses in the country with steady increase rates throughout the last decade(s) [18] Similar figures have been reported internationally. The 2009-2010 AHP Equine Industry Survey states that 78% of horse owners in the US expected to increase or maintain the number of horses they owned [4]. With rising popularity of private horse ownership dressage has recently seen a surge in interest from an amateur/hobbyist perspective. Similar to professional dressage, at the amateur level riders seek to improve the condition of their horse by instilling discipline and understanding. At a professional level dressage practice is often recorded using high-end video equipment (employing costly 3-D motion capture similar to that used in medical assessment [74]) and directly monitored by professional coaches. In contrast, amateurs – although often no less ambitious – typically do not have the means and resources for continuous professional monitoring and assessment of the horse's skills and dexterity, which often results in stagnation of development, frustration, and in the worst case injuries.

In this paper we describe the development and evaluation of a framework for automatic assessment of horse movements in dressage settings. Our approach employs miniature sensing platforms (full inertial measurement units – IMUs) that are inexpensive and can be unobtrusively attached to the horse, enabling direct movement recording with little effort and practically no hindrance for either horse or rider. Furthermore, our recording and analysis system is portable, which renders it universally applicable beyond dedicated (and costly) dressage arenas. By means of automated sensor data analysis techniques, our framework is able to provide direct, accurate, and objective feedback to the amateur rider. This allows them to gain awareness of the actions of their horse and of how they ride together, without the requirement of expensive equipment and external feedback. As such our approach is directly accessible for riders at

Level	Description
1 – Introductory	Basic gaits, and turns.
2 – Preliminary	Develops skills, training and musculature to perform the advanced level movements .
3 – Novice	Improves suppleness, balance and throughness. Introduces 15m circles.
4 – Elementary	Introduces collected work. More critical judgement.
5 – Medium	Determines the horses ability to perform medium and extended paces.
6 – Advanced Med.	Increases complexity of movements including zig-zags and five loop serpentines.
7 – Advanced	Introduces walking half-pirouettes, multiple flying-lead changes and canter quarter pirouettes.
8 – Prix St. Georges	Riders are expected to show distinct differences within the gaits from collection to extension.
9 – Intermed. I	Mental and physical preparation for Intermediate II.
10 – Intermed. II	Develops the horse for the advanced skills needed for Grand Prix.
11 – Grand Prix	The most advanced and complicated movements must be performed with absolute attention to detail.

Table 2.1: Overview of the dressage assessment procedures with the eleven levels of competitions.

all levels.

The contribution of this paper is three-fold: *i)* We describe the principles of dressage riding and judgment, and, based on this, specify a framework for automated skill assessment that is based on minimal alteration of established training procedures at the amateur/hobbyist level. We focus on optimising the level of detail that can be provided for automated quality feedback using an unobtrusive and inexpensive recording and analysis approach that allows amateur riders to effectively review their training and improve accordingly. *ii)* We present the movement recording and analysis system that we developed based on the specifications in *i)*. The system revolves around the use of wireless inertial measurement units (IMUs) [9] strapped to each of the horse's legs to measure acceleration and orientation of the horse's legs in a performance setting. Gait is classified and relevant performance metrics are extracted describing attributes relating to various aspects of the training scale, and can be interpreted by riders to identify areas for improvement. *iii)* We evaluate the applicability and the practical value of our automated assessment system in a large-scale deployment with 21 riders

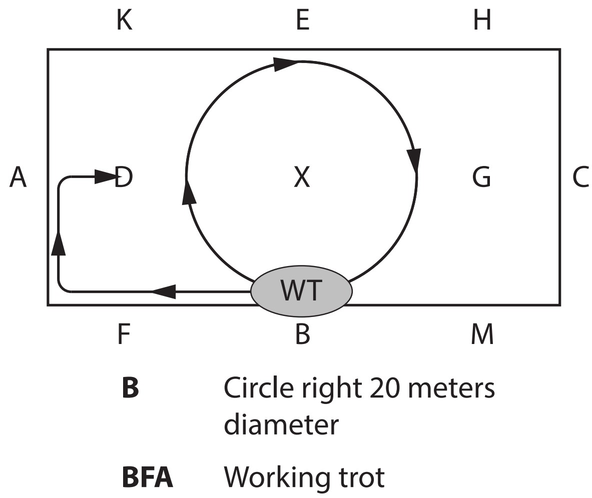


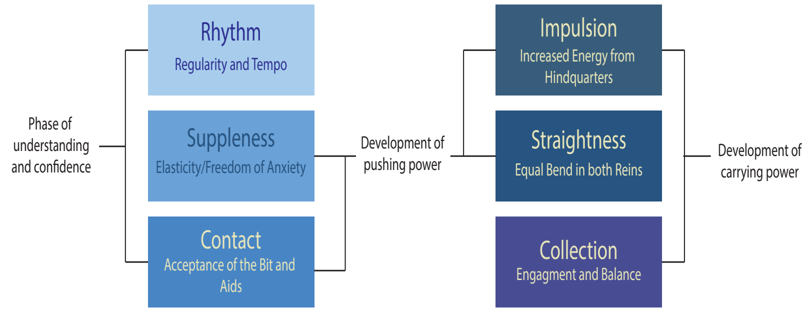
Figure 2.1: An example of a dressage movement. Representing a birds-eye view of the dressage arena, the letters around the circumference represent physical plaques that act as an indication of location for the rider. The solid line with arrows represents the desired path of the horse, and the letters in the ellipse indicate a gait transition, in this case to a working trot.

and 23 horses being assessed in standard dressage exercises. We demonstrate the practical value of the developed framework for automatically tracking key performance attributes that are of relevance for providing objective feedback on the quality of dressage movements. We show that the system produces precise feedback as we extract key, objective attributes from the performance, enhancing a rider's ability to learn from their mistakes. As such the developed framework has the potential to be of substantial practical value for the large population of dressage hobbyists.

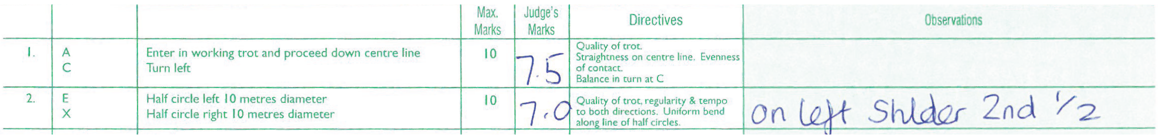
2.2 Dressage and Judgment

The horse's natural speed and endurance are a result of their role as a prey species within the ecology of the North American prairies, supplemented by a significant amount of selective breeding and training by humans [62]. Through development and conditioning it has been possible to train horses to display tremendous precision and dexterity, allowing these half ton animals to move with the utmost grace across an arena. No differently from human athletes, a horse's body and its condition are crucial to reaching the highest levels of performance. Training induces physiological adaptations that allow the horse to perform at a high level with minimal risk of injury

(a) *Piaffe* – photo courtesy of North Kaludah Dressage Horses.



(b) The training scale of dressage describing the three main phases of training: understanding, pushing power, and carrying power (from [46]).



			Max. Marks	Judge's Marks	Directives	Observations
1.	A C	Enter in working trot and proceed down centre line Turn left	10	7.5	Quality of trot. Straightness on centre line. Evenness of contact. Balance in turn at C	
2.	E X	Half circle left 10 metres diameter Half circle right 10 metres diameter	10	7.0	Quality of trot, regularity & tempo to both directions. Uniform bend along line of half circles.	on left Shlder 2nd ½

(c) Dressage score sheet.

Figure 2.2: Skill assessment in professional level dressage. (a): Rider and horse demonstrating a *piaffe*, an advanced dressage movement involving raising diagonal pairs of legs alternately whilst staying in the same place, requiring intense strength, coordination and discipline from the horse. (b): As the horse gains experience and competence at each level of the training scale attention can be paid to the next, and the horse can advance through the levels of dressage. Taken as a whole these skills contribute to the overall throughness ("Durchlässigkeit" [46]) of a horse. (c): British Dressage score sheet, completed by a qualified judge.

[62]. During dressage tests it has been shown that the horse reaches heart rates of up to 141bpm [26]; in comparison to the natural resting heart rate of approximately 36bpm, this demonstrates the exertion involved. It has been shown that it is possible to change the physiological characteristics of the dressage horse's motion through training.

Whether ridden competitively or for pleasure, dressage has been shown to have physical benefits for both the riders and the horses [62]. As a sport, dressage is performed through the execution of "tests": predefined sequences of movements of varying difficulty linked together with transitionary movements. Depending on the experience and ability of the horse and rider, different levels of dressage are performed. These range from "Introductory", the most rudimentary of exercises, to "Grand Prix", the pinnacle of the sport and the most involved and complicated exercises. The full range of levels is shown in Table 2.1. An example of a movement specification as it would be presented to a rider is shown in Figure 2.1. This represents a top down view of the dressage arena and the horse's expected path. The letters around the circumference of the arena are representations of plaques placed around the arena that give signal to the rider regarding their location. Instructions are presented in terms of locations at which movements should start and finish, and where key transitions should be made.

The skills required by a horse to perform dressage movements vary depending on the movement. Amongst the various basic gaits – walk, trot, and canter – regardless of the level of collection or extension required, regularity of pace and rhythm are key, as well as an ability to remain relaxed and concentrated under the pressure of performing. In turns and corners conformation to a specified line must be shown across the entire length of the horse's body. More complicated movements demand a commensurate level of skill to perform. A movement epitomising the dexterity and control required by the performing horse is the "piaffe", a trot in place, in which alternating diagonal pairs of legs are raised off the ground (Figure 2.2a). This requires intense training and great amounts of strength and control on behalf of the horse, but should, nevertheless, appear effortless [110]. Progress through the levels of dressage requires conformity to the "Training Scale". As shown in Figure 2.2b, there are six discrete principles of dressage which together describe the "throughness", or "Durchlässigkeit" (German; official dressage term for *throughness*) of the horse.

Throughness is a core component of judging quality in dressage, and is extremely difficult to quantify. It can be described as the connection from the bit to the hind leg, with the horse accepting the contact with submission and relaxation. The FEI judge's handbook [46] identifies the horse's confidence and understanding of the riders intention as fundamental aspects of *throughness*. The handbook suggests that these aspects can be be developed by training in rhythm, suppleness, and contact. Figure 2.2c depicts an extract from a scoresheet as it is used in official British Dressage competitions.

At a competitive level, the quality of a horse's performance is traditionally determined by a judge. During competition, a qualified judge conducts an assessment based on a set of rules provided by the FEI (International Federation of Equestrian Sports), and informed by experience, intuition and ideals of grace and elegance of form. Through

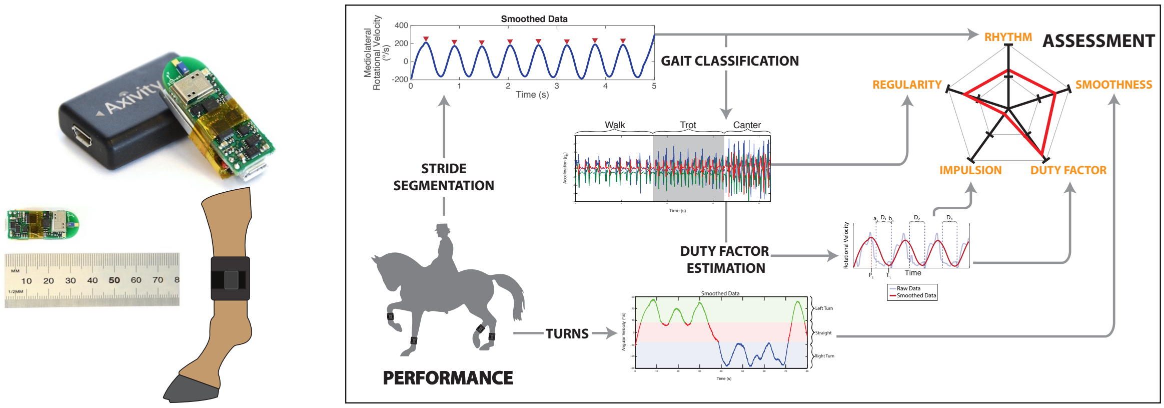


Figure 2.3: Overview of the sensing and analysis framework for automated generation of quality feedback for dressage riders. Left: Miniature sensing platforms (IMUs; Axivity's WAX9 [9]) are incorporated into a set of brushing boots. The boots are strapped around the horses cannon bone, above the fetlock. The boots used in the study were made of neoprene with velcro fastenings, meaning they were flexible and comfortable whilst maintaining a secure position. Right: Analysis workflow that measures five distinct performance attributes as they are of relevance for dressage assessment (see text for description).

practice and routine, it is possible for a rider to develop an intuition for their, and their horse's, performance. However, for the majority of amateur riders, when not in a competition, feedback is provided by a coach or trainer. These professionally qualified individuals, often riders themselves, are able to provide feedback based on their experience and understanding of the sport. The constraints of time and money, however, place restrictions on the amount of formal coaching that an individual can receive.

2.3 Automated Quality Feedback for Dressage

It is the goal of our work to develop a framework for objective quality feedback for hobby dressage riders that is: *i)* ubiquitously accessible, i.e., does not rely on substantial hardware installations (as it is the case, e.g., for high-end 3-D motion capture [144]); *ii)* easy to deploy and to use even by (technically) lay users; and *iii)* that produces understandable, i.e., intelligible and accurate assessments, which provide dedicated feedback that the rider can use in order to improve their handling of the horse and thus to improve the dressage performance of horse and rider. Figure 2.3 gives an overview of our recording and assessment framework.

From the outset of our development we discarded camera based sensing solutions even though video seems to be a suitable sensing modality. Reasons for this include large installation costs, limitations to indoor scenarios, and problems with occlusions that render camera-based approaches impractical for hobbyist use. Rather, we opted for a direct movement recording approach in which we strap miniature inertial measurement units (IMUs, consisting of tri-axial accelerometers, gyroscopes, and magnetometers) to the shins of the horses. It is worth mentioning that the horses are used to wearing shin-protectors in order to prevent injuries. Consequently, it is straightforward and comes with very little extra effort to secure the sensing platforms to the horse's legs (Figure 2.3, we integrated the IMUs into a set of ordinary and regularly used brushing boots). We chose the four limbs for movement sensing as their characteristic movements are relevant for the majority of dressage tests. We use off-the-shelf IMUs (Axivity WAX9 [9]), which are inexpensive, robust, and come with a very small form factor (Figure 2.3 – left). These four IMUs stream their sensor data to a smartphone, which the rider carries in their pocket or on an armband, that is equipped with a bespoke application capable of synchronising and storing the raw sensor data. Subsequent movement analysis is currently performed offline, i.e., after downloading the data from the smartphone. This is, however, solely because at this stage of our research we focus on the fundamental system development.

Automated assessment of dressage movements is based on the measurement of *five* fundamental performance attributes: *i)* Rhythm; *ii)* Regularity; *iii)* Impulsion; *iv)* Consistency of duty factor; and *v)* Smoothness of turns and straights. These attributes mirror guidance provided to professional judges on how they should assess the quality of dressage movements (according to the FEI official judge's handbook [46]). Whilst many of the judging attributes are somewhat subjective, we have extracted core, objective aspects of each from measurements as described below. It is our goal to provide objective feedback to the rider and as such we concentrate on aforementioned quality parameters. The above terms are described in more detail and in the context of our system in the section titled "Identifying Performance Attributes" below.

Our framework follows a linear work flow (Figure 2.3 – right). The horse's strides are automatically segmented across the whole session. This segmentation forms the basis for several of our performance attributes, as it describes the first tier on the training scale at a very low level. Stride segmentation also provides the foundation for the subsequent gait classification. Assessment of performance attributes is then pursued per movement (definition given below) and results are visualised in the form of spider plots as depicted in the figure.

In what follows we will describe the performance attributes we have identified and the technical procedures for measuring them. The basis for these are synchronised streams of data from the three individual sensors incorporated into each of the four IMUs used. As each sensor (accelerometer, gyroscope, and magnetometer) measures on 3 axes this results in 36-dimensional input data. In the course of the study, however, we determined that in order to remain useful the magnetometer would require more calibration than it was reasonable to expect a layman to perform, and, as a result, the magnetometer data was not used. Subsequently all analysis was performed on 24-D sensor data recorded with a sampling rate of 80Hz (as supported by the sensors per default).

For most of the measurement procedures described below, we use a sliding window approach that analyses sensor data in their temporal context. This is motivated by the inherent sequential nature of the data for which a singular treatment is typically not insightful. Based on informal cross-validation experiments we optimised the window length to 3 seconds, i.e., with aforementioned sampling rate analysing 240 consecutive sensor readings, as this allows the slowest typical gait – walk – to cycle fully twice and thus to capture all relevant characteristics in one analysis frame (for a horse of average build). Subsequent analysis frames overlap by 90% in order to maintain a suitable temporal resolution of the extracted attributes. In the remainder we will refer to the horses' legs by FR, FL, BR, and BL, where F and B indicate fore and back legs, and L and R indicate left and right.

2.3.1 Pre-processing: segmentation

Our dressage assessment system operates directly on the raw IMU sensor data in the sense that we measure dedicated performance attributes (as described below) for feedback generation. Effective calculation of these measures requires some elementary pre-processing steps: *i)* stride delimitation; *ii)* gait classification; and *iii)* movement segmentation. Before describing the measurement of the actual performance attributes we will first summarise these pre-processing steps.

Stride Delimitation We segment individual strides – per limb – based on peak-detection within the smoothed mediolateral axis of the gyroscope. The mediolateral axis refers to the axis that runs from the centre of the horse to the flank parallel to the floor, see Figure 2.5. When considered in the context of the sensor's locations on the leg, the gyroscope data in this axis describes the leg's swing forwards and back as each stride is made. The peaks in this data stream describe the points at which the rotational velocity is largest, that is, when the leg is swinging forward fastest. This occurs at a consistent point in each stride and consequently can be used as a stride delimiter. Noise in the signal can cause confusion when using peak finding algorithms. Accordingly, we use a lowpass Butterworth filter in order to remove the majority of the noise, followed by a Savitzky-Golay filter to clean up the resulting signal. Figure 2.4 illustrates the typical gyroscope signal for a single sensor during a trot with clear peaks per stride. Information about strides (and sequences thereof) are required for measuring a number of the performance attributes as described below.

Gait classification The execution of preliminary level dressage tests requires demonstration of the three main gaits: walk; trot; and canter. In order to assess these gaits using the quality parameters explained below, an automatic analysis system needs to classify gaits accordingly and with high reliability. In general the gait of a quadruped is – by nature – very regular and due to the complexity of coordinating four limbs there are quite significant differences between different gaits, which is somewhat in contrast to human gait. It allows us to perform gait classification solely based on timing analysis using simple thresholding methods. Without practical limitation we restrict gait classification to a horse's left foreleg.

Movement Segmentation Strictly speaking there is no technical need for limiting our measurements to certain dedicated and fixed intervals as we could calculate and

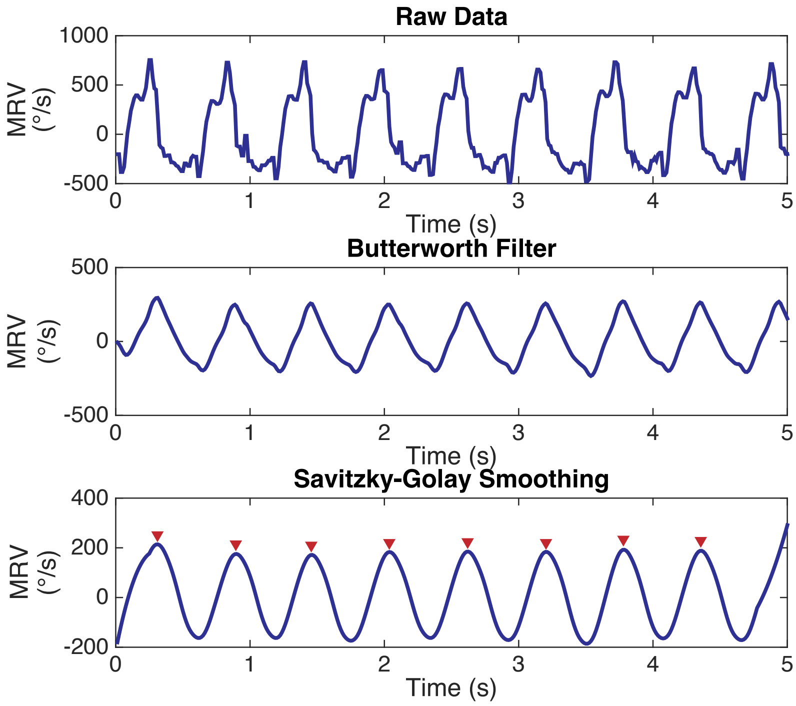


Figure 2.4: The stages of noise removal as mediolateral gyroscope data is preprocessed for stride delimitation. The red arrows in the final image indicate the positions of the stride delimiting peaks.

provide feedback continuously. However, in order to structure the generated feedback in a meaningful way we aggregate the assessments on a per-*movement* basis, which is according to the general practice of professional judges who score dressage performance in a similar way (cf. Figure 2.2c for an illustration of a judge scoresheet). Movements are defined as a predetermined sequence of gaits, transitions and the path the horse should follow (cf. Figure 2.1 (right)). As such, indicators for changes in movements – and hence "natural" segmentation points – are: *i)* changes in gait (e.g., from trot to canter); *ii)* changes in the horse's main orientation (e.g, when taking a turn); and *iii)* halting points. Technically, movement segmentation is straightforward and can be implemented with simple heuristics based on the aforementioned pre-processing.

2.3.2 *Identifying performance attributes*

As illustrated in Figure 2.2b, according to professional standards, dressage judgment is essentially based on the measurement of six fundamental aspects: *i)* Rhythm; *ii)* Suppleness; *iii)* Contact; *iv)* Impulsion; *v)* Straightness; and *vi)* Collection. These

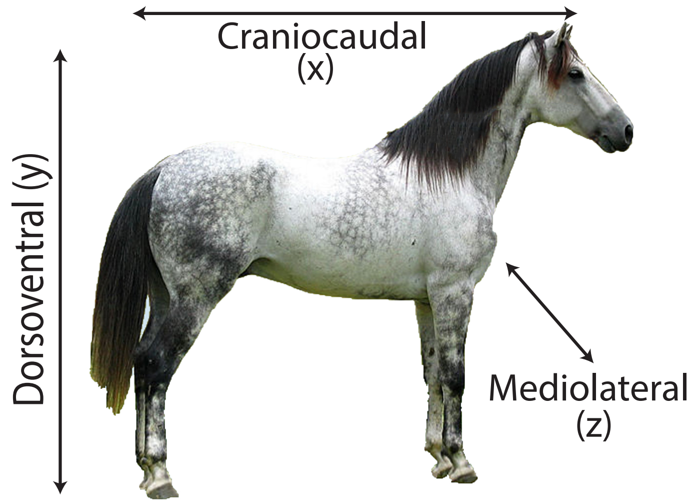


Figure 2.5: The three axes of the IMU in the context of the horse's body. The x-axis describes movement in the axis running from the head to the tail of the horse (craniocaudal). The y-axis describes movement in the axis running from the hooves to the back of the horse (dorsoventral). The z-axis describes movement in the axis running across the horse from one side to the other (mediolateral). Image adapted from original by Bill Vidigal '- CC-BY-SA 2.5

categories describe the training scale of dressage. For example, mastery of Rhythm allows for training in Suppleness. Ideally a system for feedback would provide coverage for all of the rungs on this ladder, however given the sensing modality used for this framework we have selected a few key features to focus on. These features quantify the motion of the horse's legs. The elements of the training scale that are not covered by our framework either rely on a more subjective assessment, as freedom and submission do, or refer to movement that cannot be captured by sensors on the legs. We considered the restriction of sensors to the legs to be critical to the unobtrusiveness of the system, and to minimising the imposition caused to the horse. As such we have extracted the key, measurable aspects of the training scale that are displayed through the movement of the legs and interpret them in the context of the sensed data. Below is a brief description of the performance attributes we have identified and how they relate to the judged metrics.

Rhythm The term rhythm, in this context, refers to the consistency of the beat in all paces [46]. The rate at which each of the horse's feet contact the ground should be maintained through all turns, transitions, corners and straight lines. The rate is referred to as the Tempo and is measured in beats per minute (bpm), the calculation of which is required for the calculation of the rhythm.

Regularity Each of the three major gait classes – walk, trot, and canter – has a prescribed sequence of footfalls involving all four legs of the horse that fundamentally characterises the gait. Correct conformation to this sequence is critical if a horse is to perform at any level. Deviation from these sequences – referred to as *irregularity* – during a competition would result in the attribution of an error penalty. Consequently, *regularity*, i.e., the adherence to a prescribed footfall sequence, is an important assessment parameter for dressage judgment. We choose to use the term *Regularity* to describe this attribute in this study, as we are measuring the degree of deviation from the regular gait.

Consistency of Duty Factor The duty factor of a gait is the proportion of time in which the leg is on the ground [154]. In a similar vein to the rhythm, the metric extracted from the duty factor relates to the consistency of the gait. Given that a stride can be broken into two distinct phases, the stance phase, in which the hoof is in contact with the floor, and the stride phase, in which it is not, we are able to assess the consistency of the paces with relation to these two phases in addition to the higher level rhythm. At the higher levels of dressage performance there are generally three variants of each of the three gaits, "extended", "medium", and "collected". These variants describe a wealth of subtle variation in the gaits, but can be boiled down to a continuum between an extended gait with long reaching strides in which the horse covers as much ground as possible, and the collected gait, in which the horse's strides are short and snappy, with the maximum engagement of the hindquarters. At the lower levels of dressage the horse is not expected to perform these variations, but it is expected that the gait remain consistent in terms of the extension.

Impulsion The impulsion of the horse is described in the literature as "increased energy from the hindquarters". The FEI judge's handbook notes that "the most important criteria of impulsion is the time the horse spends in the air rather than on

$$\mathbf{S}_w = \begin{array}{c c} & \begin{array}{cccc} fl & bl & fr & br \end{array} \\ \begin{array}{c} fl \\ bl \\ fr \\ br \end{array} & \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \end{array} \tag{2.1}$$

(a) Walk

$$\mathbf{S}_t = \begin{array}{c c} & \begin{array}{cccc} fl & bl & fr & br \end{array} \\ \begin{array}{c} fl \\ bl \\ fr \\ br \end{array} & \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \end{array} \tag{2.2}$$

(b) Trot

$$\mathbf{S}_{lc} = \begin{array}{c c} & \begin{array}{cccc} fl & bl & fr & br \end{array} \\ \begin{array}{c} fl \\ bl \\ fr \\ br \end{array} & \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \end{array} \tag{2.3}$$

(c) Left canter

$$\mathbf{S}_{rc} = \begin{array}{c c} & \begin{array}{cccc} fl & bl & fr & br \end{array} \\ \begin{array}{c} fl \\ bl \\ fr \\ br \end{array} & \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \end{array} \tag{2.4}$$

(d) Right canter

Figure 2.6: Substitution matrices with binary costs for regularity assessments in four relevant gait types ((a) – (d)).

the ground" [46]. This metric is only calculated for those gaits in which a phase of suspension is present i.e., trot and canter.

Smoothness of Turn and Straight A horse should execute a turn, be it a circle or a corner, smoothly and with a uniform amount of rotation for the duration of the turn. This produces the correctly shaped turns, ensuring cleanness of corners and roundness of circles and half circles. Similarly in a straight the horse's heading should remain constant right up to the beginning of the next turn.

2.3.3 Calculating performance attributes

Rhythm Rhythm and tempo T calculations as described in the following are conducted individually per limb. For overall skill assessment we then use averaged values across all four legs. Using our sliding window approach (window length of 3 seconds with a 90% overlap), the tempo T_m of each movement m is calculated as follows. Given a window w of N consecutive data samples that cover a series of $N_{\mathbf{P}}$ stride peaks $\mathbf{P} = \{P_1, P_2, ..., P_n\}$ derived using the technique described above (typically $N_{\mathbf{P}} \ll N$),

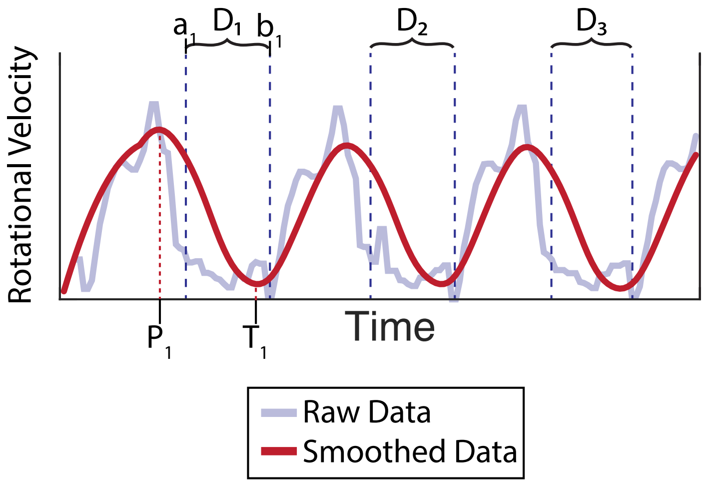


Figure 2.7: Illustration of stance phases (e.g., for estimation of consistency duty factor – see text for description).

the tempo T_w for a window w (in bpm) is calculated as:

$$T_w = \frac{\mathbf{card}(\mathbf{P})}{N * Fs * 60} = \frac{N_{\mathbf{P}}}{N * Fs * 60} \tag{2.5}$$

where Fs is the sampling rate in Hz. This calculation is repeated for every window w of the movement (total: W). It provides us with a set of W Tempo values for the movement m, which is under assessment:

$$\mathbf{T_m} = \{T_1, T_2, \ldots, T_W\} \tag{2.6}$$

where W represents the number of sliding windows covered by the movement m. The overall rhythm score R_m for the movement m is subsequently calculated as the standard deviation σ of a movement's tempo:

$$R_m = \sigma(\mathbf{T_m}) \tag{2.7}$$

Regularity An ideal sequence of footfalls is created *ideal*, of the same length N, as the session being assessed *actual*. The assessment outcome is equal to the number of steps that are out of sequence. For assessment of the walk the sequence is straightforward, as the footfalls are all simply sequential. However, for the trot and canter, gaits in which simultaneous footfalls are the ideal, a slightly different approach is required. Using a technique inspired by the substitution matrices used in bioinformatics sequence alignment [40], a substitution matrix has been built for each of the gaits.

We measure (ir-)regularity by aligning actual footfall sequences – as extracted during pre-processing using stride delimitation and gait classification – to the sequence considered ideal for the particular gaits. For a walk the ideal footfall sequence is BR – FR – BL – FL, whereas for left canter it is BL – BR & FL (simultaneously) – FR (BR – BL & FR – FL for right canter), and for trot it is FL & BR – FR & BL. We employ straightforward sequence alignment (cf, e.g., [40]) in order to detect mismatches between actual and ideal footfall sequences. We use substitution matrices with binary costs, i.e., 1 for an incorrect footfall and 0 for an acceptable footfall, allowing for reordering of simultaneous footfalls (Figure 2.6) for quantifying (ir-)regularities within a movement. For a sequence of $N_{\mathbf{P}}$ footfalls (within a movement m, previously classified as gait g) regularity $E_m(g)[0 \ldots 1]$ is defined as:

$$E_m(g) = 1 - \left(\frac{1}{N_{\mathbf{P}_m}} \sum_{i=1}^{N_{\mathbf{P}_m}} \mathbf{S}_g(\text{ideal}_i, \text{actual}_i) \right) \tag{2.8}$$

with $\mathbf{S}_g(\text{ideal}_i, \text{actual}_i)$ denoting the element of the substitution matrix according to the ideal and the actual footfall (and $N_{\mathbf{P}_m}$ the number of strides extracted for a movement m).

Consistency of Duty Factor Figure 2.7 illustrates the signal of the mediolateral axis of a gyroscope attached to a horse's leg. Raw data are depicted by the pale blue curve. The red curve is the result of our smoothing procedure. The stance phase is the period when the horses hoof is on the floor, which in the smoothed data corresponds to the period between a peak (e.g., P_1^t) and the subsequent trough (T_1^t). Through peak and trough detection in gyroscope data, which leads to a series of indices for both – $\mathbf{P}^t$ and $\mathbf{T}^t$ – we determine beginning and end points for individual stance phases D_i as follows (superscript t indicating that we are operating on indices of the signal rather than sensor values):

$$a_i = P_i^t + (T_i^t - P_i^t) * \alpha \tag{2.9}$$

$$b_i = T_i^t + (T_i^t - P_i^t) * \beta \tag{2.10}$$

where α and β are correction coefficients that take into account the different characteristics produced by the fore and hind legs to accurately demark the hoof up and down points. By performing a grid search across these parameters, values of $\alpha = 0.09$ and $\beta = 0.18$ were found to be most accurate for fore leg readings, whilst $\alpha = 0.32$ and $\beta = 0.64$ were used for hind leg readings.

Then the length of an individual stance phase D_i is determined by $b_i - a_i$. Subsequently the consistency of the length of stance (or alternatively the consistency of the duty factor) C_m is calculated as standard deviation σ of the lengths of all N stance phases in one movement (averaged over all four legs):

$$C_m = \sigma\left(\{(b_i - a_i) | i = 1 \dots N\}\right) \tag{2.11}$$

Impulsion Using the calculations described in the estimation of duty factor, i.e., essentially, through identifying the periods $FL_{up}, FR_{up}, BL_{up}, BR_{up}$ between stride peaks, we identify those periods S [in number of samples] in the gait in which all legs are suspended above the ground:

$$S = \{FL_{up} \cap FR_{up} \cap BL_{up} \cap BR_{up}\}. \tag{2.12}$$

Given that a more pronounced period of suspension is considered beneficial, we assume that a larger proportion of time during the gait in which all legs are off the ground is better. Consequently, the impulsion attribution (per movement) is given by:

$$I_m = \frac{\mathbf{card}(S)}{N_m} \tag{2.13}$$

where N_m is the number of sensor readings for the movement m being assessed.

Smoothness of Turn and Straight Given that the dorsoventral axis (Y axis on our sensors) of the gyroscope gives us the rotational velocity with which the horse is turning we infer that its variance is indicative of the consistency of the turn or straight. Defining the time-series of dorsoventral rotational velocity data $X_{leg} = \{x_1, x_2, ..., x_{N_m}\}$ for each leg, where N_m is the number of samples for the assessed movement, we define the average rotational velocity $\tilde{H}_m$ of all four legs (for a particular movement m) as follows:

$$\tilde{H}_m = \frac{1}{N_m}\sum_{i=1}^{N_m} H_m = \frac{1}{N_m}\sum_{i=1}^{N_m}(X_{\text{FL}} + X_{\text{BL}} + X_{\text{FR}} + X_{\text{BR}}) \tag{2.14}$$

Smoothness S_m is defined as standard deviation σ of H_m:

$$S_m = \sigma(H_m). \tag{2.15}$$

Summary To summarise, our assessment approach is based on the analysis of raw sensor data as they are recorded by the leg-worn sensing platforms. Stride delimitation based on peak detection in (pre-processed) gyroscope data leads to segmentation of movement data into strides. The extracted sequence of footfalls is then analysed regarding timing and sequentiality for gait classification in order to discriminate between the three main gaits: walk; trot; and canter. With this information we then extract performance attributes from all sensor data, i.e., analysing 24-D sensor streams as generated by the four IMUs worn by the horse (4 $\times$ tri-axial accelerometers and gyroscopes). Figure 2.8 gives a visual illustration of how these attributes are calculated. For every assessed movement m we derive a 5-tuple $\mathcal{Q}_m$ that quantifies the key performance attributes as defined above:

$$\mathcal{Q}_m = [R_m, E_m, C_m, I_m, S_m]^T. \tag{2.16}$$

With this quantitative representation it is now possible to perform dressage assessment and to provide direct feedback to the rider as we demonstrate in the next section.

2.4 Deployment Study

The sensing and analysis framework described in the previous sections offers a practical means for automatically generating direct quality feedback for dressage riders. This can be used for reflection and optimisation of training procedures and thus, eventually, for targeted improvements of the capabilities of rider and horse. Whilst our framework is generic and thus usable for all levels of horses and riders, we specifically focus on amateur and hobbyist level as that cohort represents a significant proportion for which, so far, detailed feedback is not as widely available as it is, for example, at the professional level where coaching and high-end recording equipment is accessible for

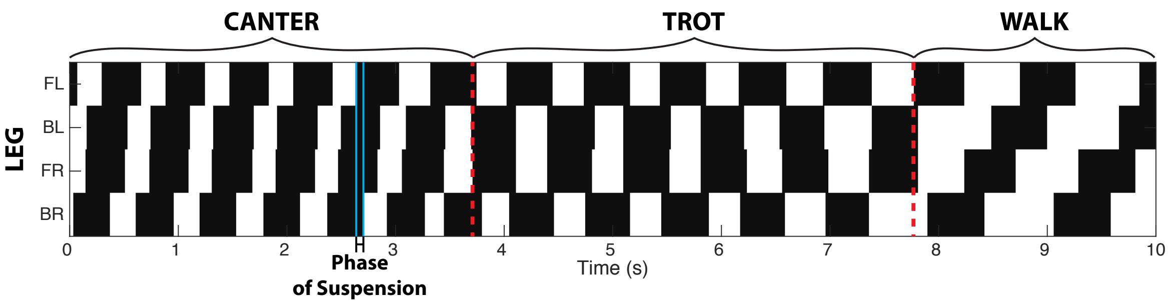


Figure 2.8: Generated automatically, this figure represents the periods when the horse's hooves are on the ground (white) and in the air (black). The data covered by the figure involves all three gaits, and the regularity can be clearly seen. It is interesting to note that the small differences in the timings of the stance phases for simultaneous landings can be observed in this figure. The *impulsion* attribute can be seen highlighted in cyan, during the canter phase (labeled as "Phase of Suspension"). This solid black region stretching from top to bottom describes the period which no hoof is in contact with the ground.

individual riders on a regular basis. In order to validate the practical value of our framework we used it in a large scale deployment where amateur riders participated in friendly competitions. In what follows we describe the details of this deployment and discuss the practical value of automatically generated quality feedback.

2.4.1 Data collection

We reached out to local horse riding clubs where dressage is regularly being practiced and recorded data for a total of 29 dressage tests that varied in the ability of the horse and rider, the difficulty level of the test and the type of horse being ridden. These tests were ridden outdoors in regular arenas, i.e., fenced areas typically used for horse riding practice. We equipped participating horses with the sensing equipment as described earlier (one WAX9 IMU for every leg of the horse, integrated into their brushing boots). Riders carried an LG Nexus 5 smartphone with our bespoke recording app running, which synchronised and stored IMU data received via bluetooth. Each test ridden was filmed to allow for subsequent annotation of the data. The data was annotated for gait, turns, and movements, by people who were familiar with the data collection process and who had an understanding of the variances of the ridden tests. The video data was distributed to several qualified judges who scored the tests performed as if in a competition setting.

Over two recording sessions we recruited a total of 23 different horses that were ridden by 21 different riders across 5 levels of dressage. The dataset contains 19 preliminary tests, and 10 tests at higher levels (cf. Figure 2.1 – left). Table 2.2 summarises the characteristics of the recorded dataset.

2.4.2 Methodology

Our technical validation focusses on two relevant aspects:

Preprocessing Procedures We evaluate the reliability of the automated pre-processing steps that are required for the subsequent measurements of the five performance attributes as described in the previous section. Specifically we report quantitative results – in terms of F_1 scores that resemble precision and recall values as typically required for such assessments [122] – for stride segmentation, gait classification. Note that movement segmentation can be considered straightforward as it is based on pre-defined rules and a quantitative evaluation would simply replicate these very rules and we thus do not show results here.

Performance Attributes The main goal of our work is to provide automated quality feedback to dressage riders. The key for an objective system of feedback is the five performance attributes we extract automatically from movement data of the horse: rhythm, regularity, consistency of duty factor, impulsion, and smoothness in turn and straight. We visualise these key attributes – and the development of their values throughout the course of an assessed dressage test – using normalised spider plots. These provide direct overviews of all five parameters in an easily graspable way and thus give direct access to quality information. Furthermore, they allow us to compare dressage tests at various levels of granularity, e.g., at the full test scale or at the level of individual (sub-) movements. We analyse the whole of the dataset recorded during our deployment study using these 5-D spider plots.

2.4.3 Results

Our performance attributes, extracted across whole tests, were used as features to train a classifier to predict the level of the performance. Using a SVM-based classification backend (with RBF kernel, and parameters optimised using standard grid search) technique [153], we were able to predict the level of the test performed (Preliminary, Novice, Elementary etc.) from our attributes with $> 75\%$ accuracy. The class weighted F-score was 0.76, with a class weighted precision of 0.77 and recall of 0.75. The confusion matrix describing the output can be seen in Figure 2.9.

This gives an indication that the metrics are discriminative with relation to the level of the horse. Classification for each gait – *walk*, *trot*, and *canter* – produced F-Measures of 0.95, 0.95 and 0.92 respectively. The weighted F-measure for all classes was 0.94.

To visualise our attributes in a way that would allow for direct comparison, we created normalised spider plots, in which smaller areas under the lines represented better assessments. In order to compare horses, the assessments were performed across the tests as a whole. To allow for comparison these assessments were grouped according to the tests that were run. This was aimed at reducing the amount of variation caused by some tests having a higher proportion of more difficult or easy movements. The resulting visualisations can be see in Figure 2.10. These visualisations give an indication of the variability of the assessments, between the performances, within the tests. We extracted descriptive statistics relating to the range of scores for each of these tests. The two Preliminary tests show a higher variability in the Rhythm attribute, with a standard deviation of 0.33 and 0.47 respectively, compared to 0.14 and 0.23 for Novice and Elementary respectively. Similarly for the smoothness of turns and straights, the Elementary level riders were more consistent with a standard deviation of 0.046 compared to the lower levels that were all over 0.06. This describes the trend that through the levels of dressage the standard of the horses becomes more consistent. As horses and riders improve past this level, more inconsistent pairs do not progress, reducing the variability within the scores. At the very low preliminary level, a vast range of abilities are present, highlighting the nature of the enthusiastic amateur.

Table 2.2: Overview of the dataset recorded during deployment study.

demographics			
# riders		# horses	# tests ridden
21		23	29
data			
total: 29 tests			
total dur.: 2.12h / avg. dur. (std.): 4.37m (±35.3s))			
#prel. 4	#prel. 18	#novice	#elementary
10	9	4	6
44.3m	33.8m	18.1m	30.8m

2.5 Related Work

Human activity recognition based on the analysis of inertial measurement data has traditionally been one of the core concerns of research in the fields of ubiquitous and wearable computing. It has typically focused on the automatic detection and classification of specific activities a person pursues in their environment [6]. A multitude of technical approaches have been proposed that enable the development of applications in domains as diverse as novel interaction techniques [78], situated support in smart environments [63], occupancy monitoring [77], automated health assessments [57, 121] or health care automation [7, 98, 120] to name but a few. The overall field has matured and standard approaches are now available to researchers and practitioners in the field [20].

Recently, the community has started exploring the use of ubiquitous computing techniques for animal related application domains. Mancini *et al.* explored the design space of human-animal interaction from an ethnographical angle [92] and made concrete suggestions for leveraging sensing and data analysis techniques for future smart environments for animals [93]. Novel interaction approaches have also been the focus of wearable sensing approaches for (rescue) dogs [64]. The foundations for activity recognition in dogs have been laid in [82], where wearable accelerometers and machine learning techniques have been used for continuous logging of dogs' activities. Within the context of activity recognition for horses only very little work has been conducted so far that uses direct sensing and automated data analysis techniques, e.g., [118]

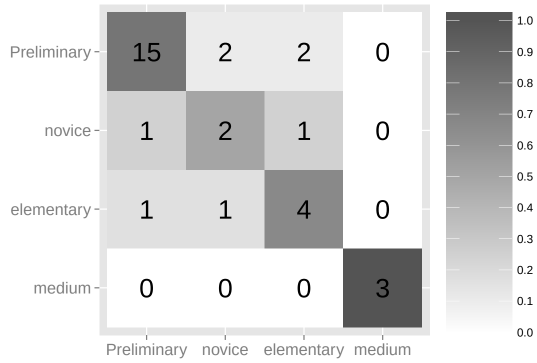


Figure 2.9: Confusion matrix showing the performance of the SVM classifier for discriminating between test level. The feature set comprised of performance attributes extracted across full tests.

where HMMs were used for stride classification.

Beyond low level classification of activities, and so far solely focusing on humans, there has been little work that focusses on quality analysis. MusicJacket [142], for example, is a wearable system that aims to support the teaching of posture and technique to novice violin players. In a similar vein, Ahmadi *et al.* [2] assess accelerometry from the swing of a novice tennis player and, through comparisons with an elite athlete's swing, are able to make inferences about the parameters of the swing that represent the skill. Other examples from the sports domain include the assessment of rehabilitation exercises [105] or the SwimMaster system [10], which analyses IMU data to evaluate the efficiency of swimming strokes. ClimbAX [83] is a system developed to automatically assess the performance of rock climbers based on data collected from accelerometers. Through identification of attributes core to rock climbing, parameters were extracted from the data that informed the user of their strengths and weaknesses.

Predominant approaches for semi-automatic analysis of the way a horse moves are assessments of video recordings for the horse (e.g., for lameness detection and analysis [69, 106, 107]). Here recording settings are typically very constrained in order to allow for robust extraction of the kinematics of the horse, which limits the wider applicability to very specific settings. At a professional level, dressage is – on very rare occasions – assessed using high-resolution 3D motion capture. There is a body of work investigating the use of accelerometers to detect lameness in livestock, including horses [24, 70, 71]. To the best of our knowledge, however, our framework is the first that focusses on automated and objective assessment of dressage exercises using wearable and thus universally applicable sensing, and sensor data analysis.

2.6 Conclusions

As with any attempt to quantify an inherently complex assessment there are bound to be limitations to the system. Throughout the course of this study we considered accessibility and unobtrusiveness of the system to be key. For this reason we restricted the sensing platform to 4 sensors that could be incorporated into a piece of equipment that the horse would have a prior familiarity with, the brushing boots, see Figure 2.3. With a more comprehensive sensing system, a more detailed picture of the horses' movements could have been recorded. This would have given access to assessments of the horses form aside from the movement of the legs, however was seen as too much of an imposition on an amateur rider and horse to maintain wide ranging applicability.

Another implication of our sensing platform was that spatial information about the performance was not available. Positioning within the arena is an important factor of judging, however attaining information of the precision required from IMUs is not currently feasible. Even with advanced GPS chips the spatial resolution is in the order of meters. Given that some movements require consecutive changes of direction within short space, this would not have been adequate.

2.6.1 Summary

Dressage is popular both from a competitive and hobbyist perspective. In the course of our study it was apparent just how many horse owners would take part in small, unaffiliated competitions for the opportunity to get feedback on both their horse and their riding. Like many sports, progress is dependant on deliberate practice, and as such entry into competitions or the hiring of a coach is an important factor in any dressage rider's training. However, these avenues for feedback are costly, both in terms of time and money, which thus often cuts out hobbyists from detailed and quality feedback.

We have developed a sensing and analysis framework that automatically generates quality feedback for dressage exercises. It is based on miniaturised inertial measurement units (IMUs), which are integrated into the standard brushing boots of a horse. IMU data are streamed via bluetooth to a smart phone in the rider's pocket. This sensing approach allows for inexpensive, portable and simple data collection. Using an automated pre-processing and analysis approach our framework generates reliable and reproducible measurements of key performance attributes that are of relevance for describing the characteristics of dressage movements. We use these performance attributes for quality feedback to the riders at varying levels of granularity.

We deployed our framework in a large scale study that involved 21 riders and 23 horses. For a total of 29 dressage tests, performed at varying levels of complexity we generated quality feedback based on the automatically measured performance attributes. The automated feedback system is able to provide specific insights on the quality of the ridden dressage tests and to indicate where and how a rider could improve their performance.

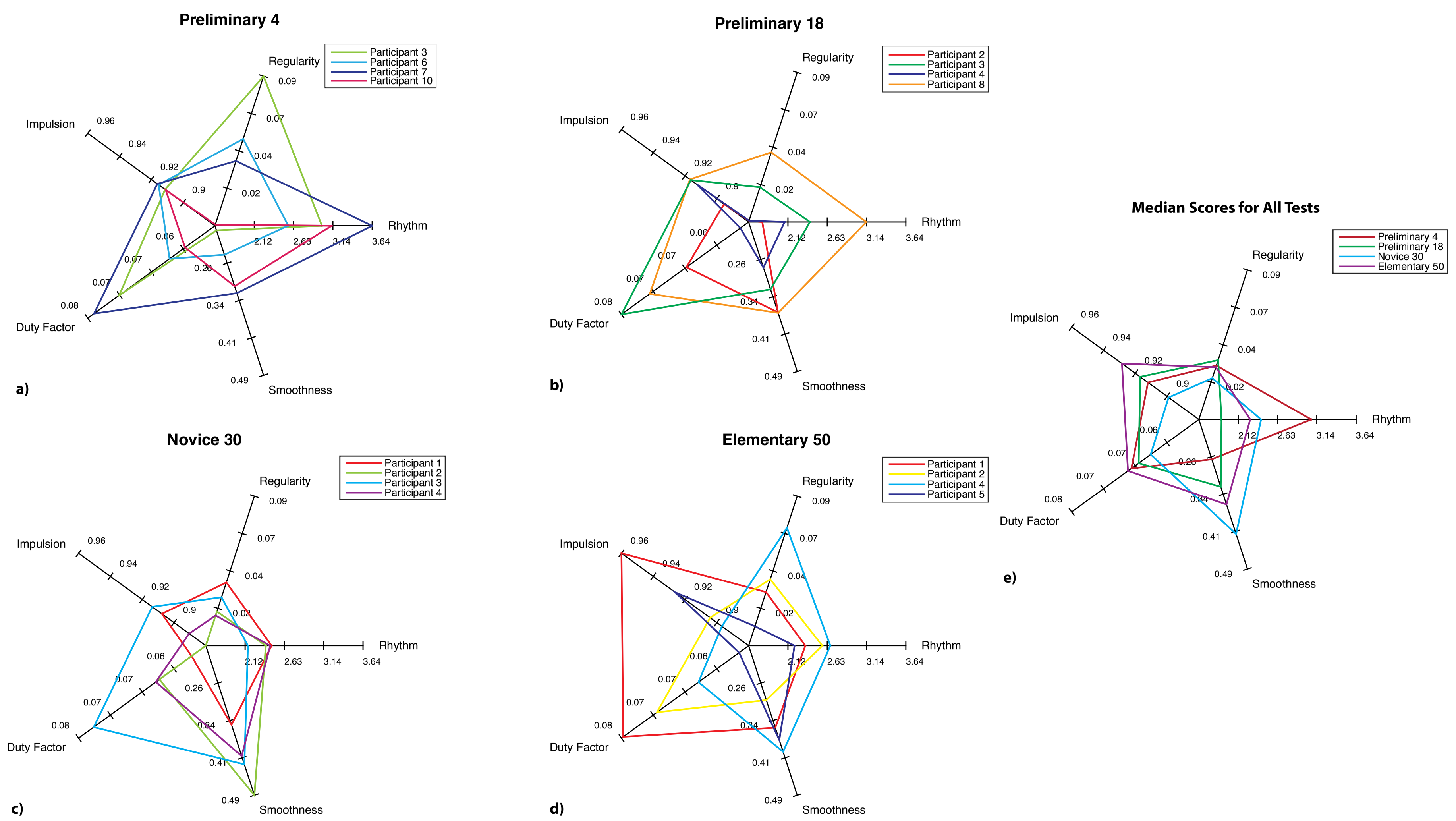


Figure 2.10: Automatically generated performance attributes for dressage tests (selection from deployment). Smaller values correspond to better performance. For objective rider comparability attributes are grouped according to tests assessed with increasing level of complexity for the overall tests: a) Preliminary 4 (easiest); b) Preliminary 18; c) Novice 30; d) Elementary 50 (hardest; cf. Figure 2.1 for explanation of test characteristics). a) to d) show per participant scores. 4 participants have been selected randomly to reduce visual clutter in the spider plots. Considerable differences in performance can be seen, e.g., participant 4 in Preliminary 18 performing better than Participant 8. e) shows median values for each of the test groups, which illustrates how the challenges to the horse and rider generally vary between the different tests. See text for details, best viewed in colour.

3

Porcine Lie Detectors

Automatic quantification of posture state and transitions in sows using inertial sensors

Contents

Abstract

Automatic classification and quantification of postures and posture changes in domestic animals has substantial potential to enhance their welfare and productivity. Analysis of such behaviours in farrowing sows can highlight the need for human intervention or lead to the prediction of movement patterns that are potentially dangerous for their piglets, such as crushing when lying down. We present a novel approach to automated classification and quantification of sow postures and posture transitions that enables large scale and accurate continuous behaviour assessment on farm. We use data recorded using a tri-axial accelerometer attached to the hind-end of the sow, from a deployment that involved six sows over the period around parturition. We automatically classify the posture state – such as standing, sitting, lateral and sternal lying – of the sows for the full dataset with a mean F_1 score (a measure of predictive performance between 0 and 1) of 0.78. Posture transitions were detected with an F_1 score of 0.79. We automatically extract and visualise a range of features that characterise the manner in which the sows change posture in order to provide comparative descriptors of sow activity and lying style that can be used to assess the influence of genetics or housing design. Through automated detection of changes in activity we are able to predict events such as the onset of farrowing. The methodology presented in this paper can be readily applied in large scale deployments with substantial potential for enhancing animal welfare and productivity on farm.

3.1 Introduction

Automatic classification and quantification of postures and posture changes in domestic animals has substantial potential to enhance their welfare and productivity. Freedom of movement was one of the original “Five Freedoms” in the seminal Brambell Report on farm animal welfare [17] and the ability of housing systems to deliver this in a species-relevant way is a key component in modern welfare assessment schemes [15]. Changes in posture and activity may also be indicative of impending health problems [150]. In the case of the domestic sow, automated posture assessments may facilitate the identification of additional specific behaviour traits that may confer advantages or disadvantages to the production system. Detection and analysis of activity patterns preceding farrowing may indicate the timing of parturition and the need for human intervention. Since the sow poses a significant crushing risk to the piglets [95, 119, 130, 151], the way in which she lies whilst in farrowing accommodation relates to her maternal ability and the adequacy of the housing provision, and has consequences towards the survival of her piglets. Selection in swine production has resulted in a change of body shape leading to changes in the amount of control a sow can exhibit during lying [94]. Consequently, many piglets are crushed either as the sow lies down (standing to lying event) or when she rolls from lateral side to lateral side (rolling event). As an approach to accounting for this, the prevalent housing system for farrowing sows confines them to farrowing crates that restrict their movements, increasing survival rates of piglets by effectively minimising the risk from crushing [35]. Given that there is substantial variation in sow activity and lying behaviour [95, 119, 130, 151] between individuals, categorisation of sows according to their lying behaviour, and the selection of breeding lines based on this, could be used in addition to selection of most appropriate housing design and management to improve the welfare and productivity of the farrowing system. With several sows farrowing at any one time in typical pig units, and potentially large numbers of animals required to perform genetic selection for aforementioned traits [79], automated assessment methods are a necessity for large scale utilisation of quantitative posture information. The objective of this paper is to describe a methodology for automated posture assessment in sows around and during parturition. A Tri-axial accelerometer, attached to the monitored animal is used to collect relevant data. Accelerometers have previously been used to detect oestrus [29], sow posture [31] and sow activity before, during and after farrowing [30, 111], as well as the onset of parturition in sows [33]. The novelty of our work lies in:

1. The placement of the sensor, which allows the collection of more relevant data on posture transitions.

2. Specifically targeting posture transitions for analysis.

3. Focusing on the detailed analysis of sow movements, specifically targeting the

prediction of behaviours dangerous for offspring.

Prior work has revolved predominantly around posture state observations and activity levels [30–34, 111]; by considering the periods in which a sow is moving between posture states, we can make valuable assessments of the lying style. Using the methods developed in this work we aim to conduct a preliminary investigation into how a framework can be implemented to allow for detailed assessment of sow movements. Based on the conclusions drawn from this work, a further large scale deployment of the technology will be performed in order to draw conclusions targeting the prediction of dangerous behaviours and the onset and duration of farrowing; the identification and analysis of nest building activity; and overall maternal performance.

3.2 Materials and Methods

3.2.1 Data collection

Six hybrid sows from the Newcastle University Cockle Park Pig Unit, due to give birth concurrently as part of a batch, were selected for assessment and moved to farrowing accommodation five days prior to their expected farrowing. The sows had all farrowed at least once previously, being either in their second or third parity. They were housed in standard farrowing crates with a concrete floor at the front and slats at the rear. All pens contained an area for the piglets in a front corner, with a heat lamp (alternately front left or front right corner, three of each). Motion data were recorded from each sow around the expected period of parturition in order to detect posture and posture transitions occurring both before parturition, and also in the presence of piglets. Data collection was scheduled to run for four full days; however, the amount of data collected was reduced for two of the sows. A summary of the dataset can be seen in Table 3.1. The onset of parturition varied between sows.

3.2.1.1 Sensing protocol

An Axivity AX3 logging accelerometer [9] was attached to each sow using a combination of adhesive tape and glue according to a predefined protocol. All sows were

Table 3.1: Dataset description for the study. Six sows were used and in each case the intention was to record sensor data for four days. In the case of Pig072 and Pig096 the sensor did not record for the full duration, see notes for explanation.

Sow ID	Video Duration [hours]	Sensor Duration [hours]	Notes
Pig072	89	69	Sensor fell off
Pig096	95	74	Sensor failed
Pig106	95	95	
Pig107	96	96	
Pig235	96	96	
Pig252	96	96	

shaved and cleaned in a small region between the tail-head and hip bones. Attaching the sensor in the region between the tail-head and hip bones (Figure 3.1) allows specific insight into the forces associated with posture transitions that have been shown to pose the most danger to the piglets [14], particularly ‘flopping’ which involves a rapid vertical displacement of the hind-end of the sow [11]. The sensor was wrapped in duct tape and attached to the shaved area on the sow with strong double-sided carpet tape ensuring a consistent sensor orientation. A coating of Evo-stik™Instant Contact Adhesive was applied in a 2-inch area around the sensor with several strips of adhesive tape covering the glue and sensor for protection. The sensor collected motion data measured between $\pm 8g_0$ in the three spatial dimensions and sampled at $100Hz$ as standard.

3.2.1.2 Annotation

In order to provide a verifiable record of the movements of the sows independent from the sensor data, the sows were filmed continuously for approximately four consecutive days after placement in the crates. The camera was mounted above and to the rear of the pen and images were captured using Geovision™software on a PC. Due to the restrictions on movement imposed by the farrowing crate, the full range of sow motion was visible in the recording. There was a small amount of variation in the length of the videos due to the timings at which the video footage was retrieved. In the case of Pig072, a section of footage was corrupted for the final seven hours of the study. Video footage was annotated using the open source application ELAN [19]. Accelerometer

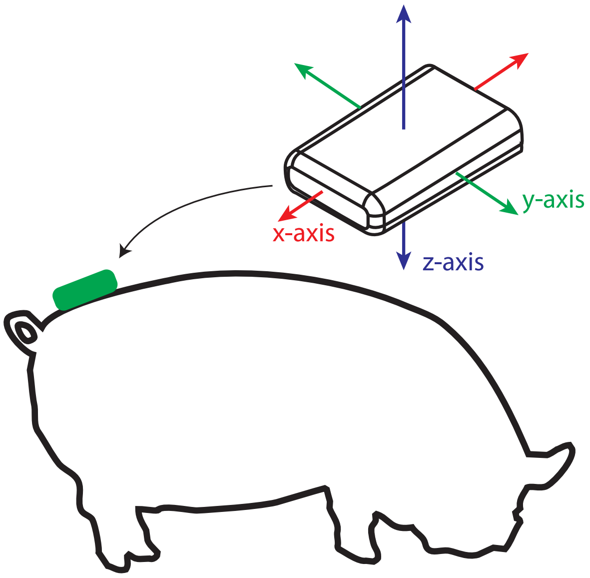


Figure 3.1: Positioning of sensor on the body of the sow. The orientation of the sensor is shown in relation to the sow. The sensor was secured to the hind-end of the sow between the hip bones and the tail-head. The x-axis of the sensor maps to the craniocaudal axis, the y-axis maps to the mediolateral axis, and the z-axis maps to the dorsoventral axis of the sow.

data were synchronised with the footage, allowing the annotations to be associated with the data directly [121]. The footage was annotated both for transitions between postures, and posture state. The annotations marked the start and end of each posture transition, and the start and end of the periods in which the sows were in a consistent posture state.

3.2.2 Detection and segmentation of posture transitions

An analysis workflow was developed in order to automatically determine the periods within the data in which the sow transitioned from one posture to another (Figure 3.2).

3.2.2.1 Preprocessing and feature extraction

Static acceleration i.e., the acceleration of the sensor due to the effect of earth's gravity, was estimated from the raw signal using a moving average filter. The output from this process was used for detection of transitions and classification of posture, limiting the effect noise has on the analysis later in the workflow. Given raw signal samples

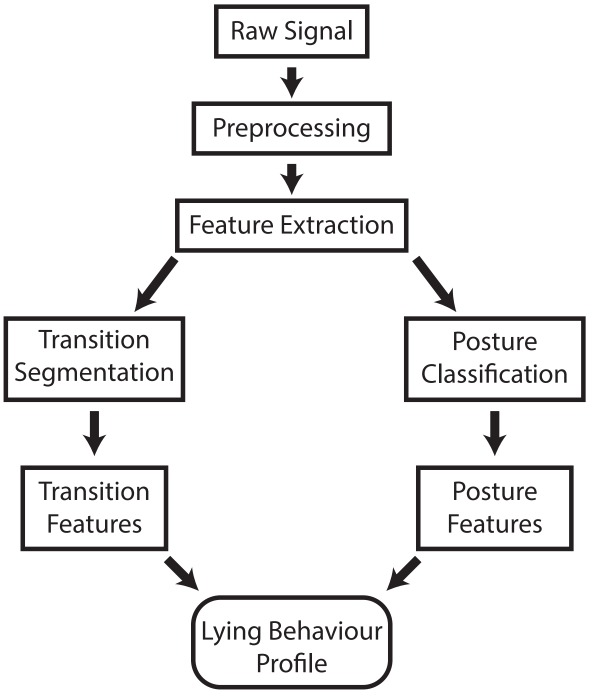


Figure 3.2: An overview of the processing workflow

$\mathbf{s}(t) \in \mathbb{R}^3$, with $-8 \leq s_i(t) \leq 8$, for each spatial axis i, the static acceleration was calculated according to Equation 3.1:

$$\mathbf{g}(t) = (g_x g_y g_z)^T \quad \text{with} \quad g_t(t) = \alpha \cdot \mathbf{s}_i(t) = (1 - \alpha) \cdot g_i(t-1); i \in \{x, y, z\} \tag{3.1}$$

where α is a weighting component used to vary the amount that the signal compensates for rapid changes in acceleration. The value for α is established through an empirical process in which the noisy component of the signal is minimised. For further analysis, the data were divided into short, continuous windows (frames) covering two seconds of sensor readings, i.e., 200 samples at $100Hz$. Three descriptors (features), summarising the data in the frame, were then extracted for each frame to describe sow posture: pitch, roll and level of activity. Pitch and roll, when calculated from static acceleration, describe the orientation of the sensor in 3-dimensional space. Pitch represents rotation of the sensor back and forth around the mediolateral axis, and changes when the pig lowers its hind-end to sit. Roll measures change in rotation of the sensor to the left and right in relation to the pig, and changes when the pig moves between lying on either side. Given alignment as shown in Figure 3.1 , with the z-axis follows the dorsoventral direction, the y-axis follows the mediolateral direction and the x-axis follows the craniocaudal direction. Using the estimation of the static acceleration $\mathbf{g}(t)$ (Equation 3.1) pitch p, and roll r were estimated as follows:

$$p = \arctan\left(\frac{g_y}{\sqrt{{g_x}^2 + {g_y}^2}}\right) \tag{3.2}$$

Where $p \in \mathbb{R} :]-90 \ldots 90[$. Pitch represents rotation around the mediolateral axis.

$$r = \arctan\left(\frac{-g_x}{g_z}\right) \tag{3.3}$$

Where $r \in \mathbb{R} :]-180 \ldots 180[$. Roll represents rotation around the craniocaudal axis. The final feature calculated from the data was the standard deviation of the magnitude of the signal (within a frame) and was used as measure of the level of activity. The time-series describing the magnitude of the signal, $\mathbf{m} = \{m1, m2, \ldots, mT\}$, where T is the number of samples in the frame, was calculated as below for each sample in the sub-frame:

$$m(t) = \sqrt{x(t)^2 + y(t)^2 + z(t)^2} \tag{3.4}$$

The level of activity for the whole frame was then calculated as:

$$a = \sigma(m) \tag{3.5}$$

Where σ denotes the standard deviation.

3.2.2.2 Transition detection and segmentation

Feature extraction results in three features for each frame of data $f(i) = \{p, r, a\}$. Posture transitions were detected by observing changes in the pitch and roll of a particular sensor that exceed empirically determined thresholds θp for pitch, and θr for roll. To provide a more robust assessment of the change in posture, each frame $f(i)$ was combined with the two subsequent frames into larger 12-second frames with an overlap of 60%.

For this new set of extended frames across the session, the frames $f(i)$ and $f(i-1)$ were considered to describe a transition point if the below rule holds:

$$|f(i)_p - f(i-1)_p| > \theta_p \quad \vee \quad |f(i)_r - f(i-1)_r| > \theta_r \tag{3.6}$$

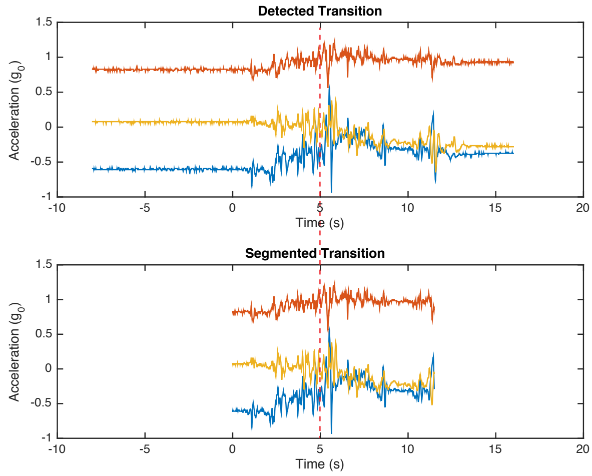


Figure 3.3: An example segmentation of a posture transition. The transition point occurs at five seconds, indicated by the dashed red line. By performing a search for stabilisation of the features either side of this point, the start and end of the transition is identified.

Frames adjacent to the transition point were also required to satisfy the above rule to ensure that a transition had occurred, rather than a temporary adjustment of position. The actual point of transition was considered as the mid-point between the frames in which the threshold is initially exceeded. Segmentation of the transitions was performed by identifying the point in the data either side of the transition at which the pitch and roll stabilise. In order to accomplish this, the change in pitch and roll was evaluated starting from the transition point. The start and end point of the transition was marked where the sensor was no longer in a consistent orientation for 3 consecutive sub-frames i.e., 1.5 seconds, to within a degree of tolerance. The result of this process is illustrated in Figure 3.3

3.2.3 Posture classification

In order to classify sow posture automatically, a Support Vector Machine (SVM) classifier using a Radial Basis Function (RBF) kernel, was trained on the frames of feature data [153]. Parameters were optimised using standard sequential minimal optimisation (SMO) and hyper-parameters were optimised using grid search. The classifier processes feature vectors extracted from the shorter, 2-second frames, and predicts class (posture) associations for each frame. A sow can be in one of five, mutually exclusive postures at any time: Standing (ST), Sitting (S), Lying on left hand side

(LL), Lying on right hand side (RL), and Sternal Lying (SL). Kneeling, a posture in which the sow rests on her front knees with her hind end raised, was considered to be a transitory posture.

3.2.4 Transition feature extraction

A set of descriptive features were identified which characterise each transition:

1. The duration of the transition described the amount of time the sow spent between posture states, and thus relates to the speed with which the transition was made.

2. The peak acceleration, as described by the maximum value of the magnitude of the signal recorded during a transition, considering all axes together.

3. The range of acceleration describes the point in the transitions when the largest change from acceleration to deceleration took place and calculated for each of the three acceleration axes. The maximum of these three is used, representing the largest change in acceleration in any one direction.

4. The rate of change of the acceleration of the sensor, known as the 'jerk'. When a sow lowers herself to the ground she may come to a stop suddenly or gently. The former will produce a large deceleration value, and similarly the jerk will be large. In the latter case, the deceleration is the same, however occurs over a longer period of time, producing a lower value for jerk. Jerk is defined formally as the derivative of acceleration.

5. The rate of change of both pitch and roll were used to define the smoothness of the transitions. In contrast to jerk however, this pair of features is calculated based on preprocessed data, rather than on the raw values. These features are calculated as the derivative of pitch and roll.

3.2.5 Lying behaviour profiles

In addition to extracting features based on the individual transitions, characteristics of the datasets relating to the lying behaviour of each sow throughout the period in which the sow was under observation were extracted. Key amongst the descriptors extracted was the quantification of how each sow preferred to lie. By collating the results of the posture classification for each sow a time budget was produced which outlines the proportion of time the sow spends in each posture. The frequency with which the sow changes its orientation was used as a robust indicator of the level of activity at a higher level than the features used for signal analysis. This was performed by taking a moving average of the number of transitions in each 2-hour period, in increments of 12 minutes.

3.2.6 Evaluation and validation

3.2.6.1 Transition detection and segmentation

The output of the detection and segmentation process is a 2-dimensional vector of timestamps for each transition indicating the start and end of the transition. An event based evaluation was performed to determine true positives (TP) false positives (FP), and false negatives (FN). The algorithm is said to have achieved a true positive detection when the detected transition coincides with a labelled transition, according to the annotated ground truth. If more than one detection coincides with the annotation, it is considered as a single correct result, however this is rare within the data. A false positive occurs where a detected transition does not lie within any annotation. Based upon these results, three standard measures for prediction evaluation were calculated: precision, recall and F_1 score. Precision describes the fraction of the transitions predicted as positive that are actually positive:

$$Precision = TP/(TP + FP) \tag{3.7}$$

Recall describes the fraction of all transitions that are predicted as positive:

$$Recall = TP/(TP + FN) \tag{3.8}$$

The F_1 score is the harmonic mean of precision and recall and reflects both in an intuitive manner:

$$F_1 = 2((recall \cdot precision)/(recall + precision)) \tag{3.9}$$

3.2.6.2 Posture classification

In order to evaluate the classifier a "leave-one-pig-out" approach was employed, in which the classifier is trained sequentially using five of the six sows, and tested on the sixth. The mean of the evaluation measures across the six experiments is taken. This gives a measure of the classifiers ability to accurately predict posture on unseen data. A confusion matrix provides an intuitive visualisation of the classification results.

3.2.6.3 Transition features and lying profile

In order to evaluate the suitability for comparison of the features selected for analysis the empirical cumulative distribution function (ECDF) of each feature was calculated. The value produced by the ECDF function $F(x)$ is equal to the proportion of transitions that produce a feature value lower than or equal to x. The shape of the distribution function is used to compare the variation of the feature distribution between sows. By calculating the ECDF of a feature using transitions from all sows, a baseline for each individual sow to be compared against is established. The EDCF of the features for each sow individually was calculated and compared with the baseline.

3.3 Results

3.3.1 Transition detection and segmentation

The dataset recorded contained a total of 1268 transitions. Of these, 965 were correctly identified (true positives), whereas 303 were not detected (false negatives). In 204 frames the data were incorrectly labelled as being a transition (false positives). The precision and recall values were 0.826 and 0.761 respectively. Considering these two results, the F_1 score was 0.792. The precision value of 0.826 demonstrates that less than one fifth of the positive predictions were incorrect. A recall value of 0.761 showed that less than a quarter of actual transitions were missed. We did not weight either precision or recall above the other, consequently the F_1 score gives a representative combination of the two metrics.

Table 3.2: F_1 results produced by the Support Vector Machine (SVM) classifier when validated using a leave-one-pig-out cross validation technique. Leave-one-pig-out cross validation refers to the process of training the classifier on data from five of the six sows and testing it on the sixth. This process is repeated leaving each sow out in turn. The classification results for each test set are then combined to produce results for the full data set.

Posture	Number of frames	F_1 score (LOPO)
Standing	70,896	0.749
Sitting	11,121	0.542
Left Lie	340,423	0.900
Right Lie	370,209	0.926
Sternal Lie	144,905	0.764
Mean		0.776

3.3.2 Posture classification

The mean F_1 score for the classifier across all classes was 0.776. A class-by-class breakdown of results is shown in Table 3.2. Figure 3.4 provides a visual description of the classifiers results. Of the five classes, F_1 scores for all but one class are over 0.74, however, the F_1 score for the sitting class is lower at 0.542. This indicated a higher degree of confusion between sitting and the other classes. The confusion matrix (Figure 3.4) shows that the misclassifications of the sitting class were mostly classified as Standing and Sternal Lie.

3.3.3 Transition feature extraction

Figure 3.5 shows the ECDF plots for the six features described in the Materials and Methods (Section 3.2) i.e., transition duration, range of acceleration, jerk, maximum acceleration, rate of change of pitch, and rate of change of roll. It can be seen, for example, that a particularly inactive sow (Pig235) produces noticeably different feature distributions for several of the features. The ECDFs are generated from the underlying distribution of the feature values which are the primary output of the algorithms. The plots can be interpreted to assess the behaviour of the sows across all transition events, and also provide information regarding the shape of the feature distributions. A good example of this is in the ECDF plot for the duration feature. It can be seen that the curve for Pig072 lies well above the baseline - which described the distribution across all

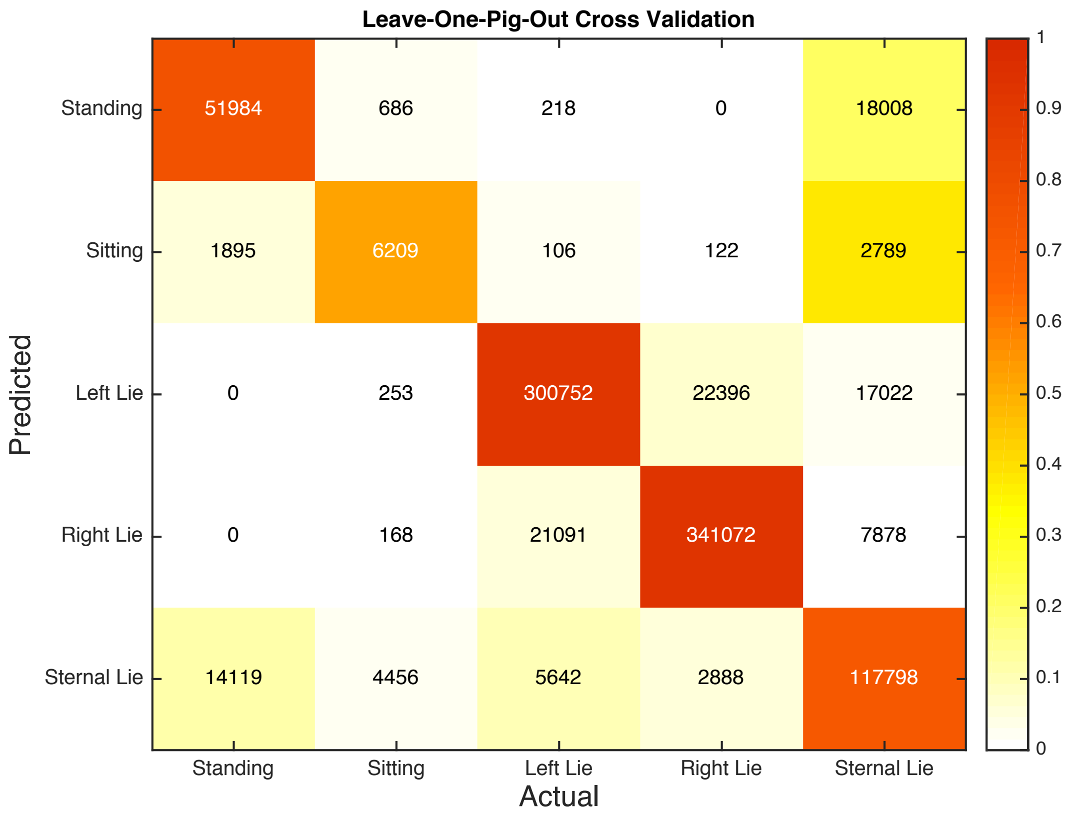


Figure 3.4: Confusion matrix outlining the results of predicting posture using a Support Vector Machine (SVM) classifier. The colour represents the proportions of the class that were predicted in relation to the number of instances of that class, a darker red is considered a better result. The number inside the cell denotes the number of instances that were classified according to the labels on the axes. A strong dark series of cells diagonally across the matrix reflects accurate classification, as these cells show the proportion of correctly predicted frames. Darker cells outside of the main diagonal describe a higher proportion of frames misclassified.

transitions for all sows - suggesting that this sow displays a larger proportion of short duration transitions. Conversely the curve for Pig107 lies underneath the baseline curve, indicating this sow displays a larger proportion of long duration transitions. Of particular interest is Pig235, which farrowed within the first 6 hours of the study. This sow has a markedly different shaped distribution for both range of acceleration and for maximum acceleration. The curve for the maximum acceleration feature lies below the baseline, indicating that the sow displays a larger proportion of transitions with a higher maximum acceleration. Whereas the curve for range of acceleration lies noticeably above the baseline curve, indicating a larger proportion of the transitions have a small range of accelerations.

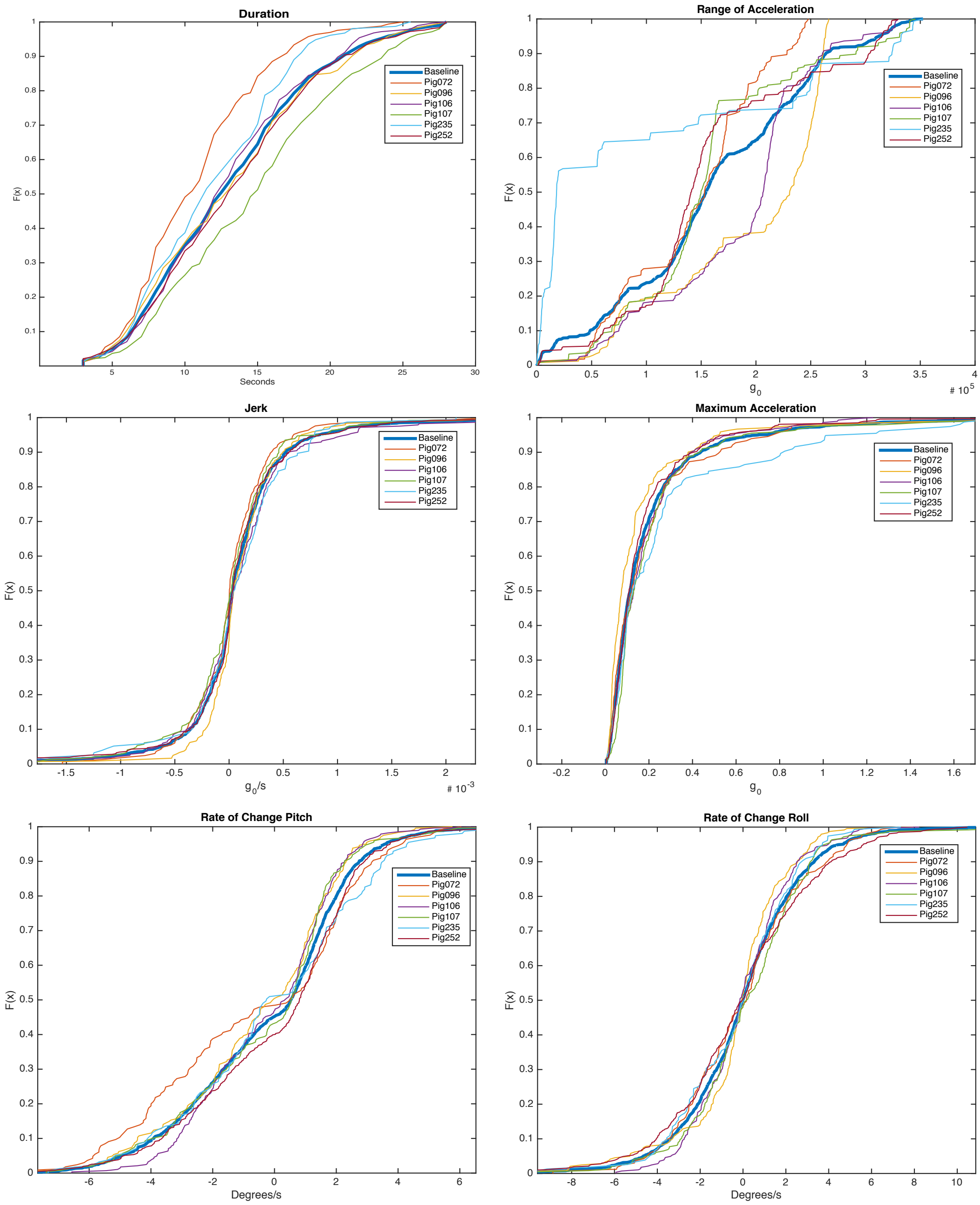


Figure 3.5: Empirical Cumulative Distribution Function (ECDF) plots for descriptive features extracted from posture transitions. ECDF plots show the cumulative proportion of instances on the vertical axis plotted against the feature value on the horizontal axis. The thick blue line describes the ECDF of all transitions across all sows. The other lines represent the ECDF plots for each individual sow. Note the cyan line in the range of acceleration from Pig235, which describes a distribution containing a transitions with smaller ranges of acceleration values, whilst the distribution for the maximum acceleration shows that larger accelerations are more common than in the other sows. Best viewed in colour.

3.3.4 Lying behaviour profiles

A visualisation of the transition frequency data can be seen in Figure 3.6. These charts provide a representation of the sows' activities throughout the recording periods.

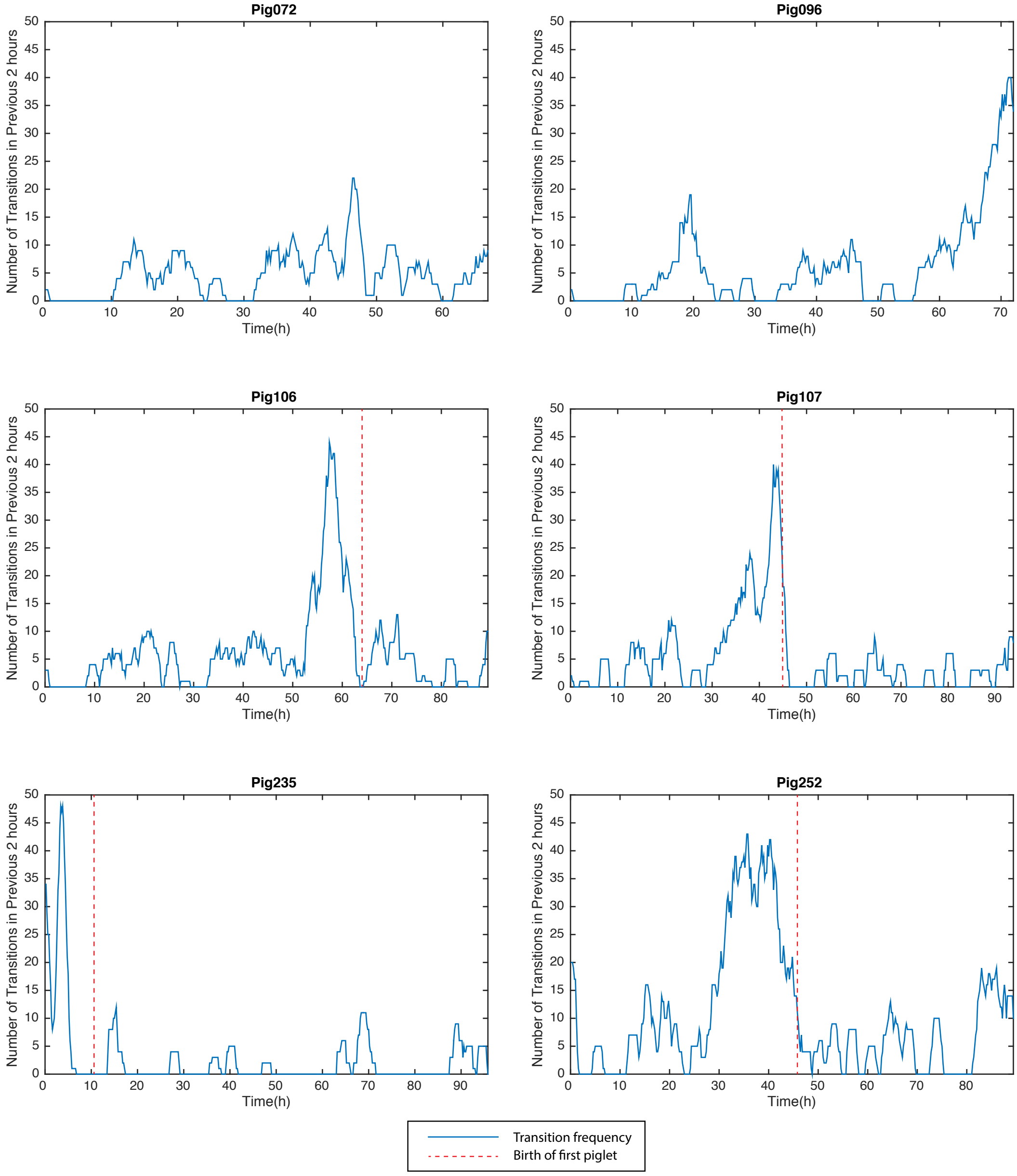


Figure 3.6: Posture change frequency. A moving average is taken over a 2-hour period in increments of 12 minutes. The large spikes shown in the charts for Pig106, Pig107, Pig235 and Pig252 coincide with the increased activity in the build up to farrowing. The dashed red line indicates the period in which the first piglet was born. Pig072 and Pig096 did not farrow whilst the sensor was recording, although Pig096 farrowed shortly after the sensor stopped.

3.4 Discussion

The objective of this paper was to develop a methodology for automatic quantification of posture and posture transitions of sows around, and during, the period of parturition. This has been achieved through analysis of data collected from a pilot study, using six sows, with over 525 hours of accelerometer data. The use of accelerometer data to record the movement of animals is well established, however in the design of this study other options were considered. “Computer vision” describes a set of techniques that enable automatic interpretation of visual data to be conducted. There are specific issues that preclude the use of computer vision in the context of this study however. Primary amongst these was the study’s aim of generating detailed representations of posture transitions and the associated forces. Without biometric measurements specific to the individual sows, it is not possible to measure force, however, insight can be gained into the forces exerted during transitions from the acceleration data as the two measurements increase proportionally. It could be argued that the inclusion of a gyroscope in the sensing platform would have been appropriate, however, due to the power requirements of a gyroscope (depending on the device, up to 20x the requirements of an accelerometer), the sensor would have been unable to record for the intended duration of the study, making it unfeasible.

Given the requirement that detailed information relating to the accelerations exhibited during posture transitions should be generated, the positioning of the sensor was a key consideration. Previous studies investigating the classification of posture and activity in farrowing sows have employed sensors attached to the animals’ crates, collars and legs [31, 124, 146]. Collar mounted sensors have limited capacity to accurately observe posture changes at the hind-end of the sows, consequently the decision was made to position the sensor on the sow’s hind-end above the tail-head and below the hip bones. This is an approach that has not been reported in the literature, and as such provides an entirely novel perspective for this kind of analysis.

The precision and recall values for the segmentation algorithm indicate that a sizeable proportion of posture transitions were correctly identified, however, these results also demonstrate that there is scope for improvement in the system. A portion of the false positive results were produced when the sensor recorded a large movement of the sow, away from its current position, in order to scratch or shift temporarily for example, before returning to its original posture without changing posture classes. Adjustments were made to the algorithm in order to correct for this however, where a sow's posture after the movement was sufficiently different to the original posture (although not in a concretely different posture category) the segmentation algorithm was prone to recording a false positive detection. Mis-detections were also produced by very gradual changes of posture over a period of several minutes. These transitions occur mainly between sternal and lateral lies, although the data set includes transitions between a sit and a lie where this also occurs. The failure to detect the transition occurs due to the orientation of the sensor consistently changing by amounts lower than the thresholds used in Equation 3.6, resulting in false negative results. Future work would involve eliminating these causes for failure. This could be achieved through the use of an additional sensor in a different location on the pig to verify posture changes, however this would necessarily increase the complexity of the system and would require a different approach.

There is a body of work that focuses on behaviour analysis of farrowing sows, and in particular the use of motion sensing platforms to perform activity recognition, although very little attention is given to assessment of posture transitions, outside of this work. For example, Cornou & Lundbye-Christensen [32], consider activity recognition in farrowing sows as a two class problem, lying vs active, and report an accuracy of up to 97%. Marchiorio et al. [97] employ an heuristic based approach to classify activity states in sows, where the states are classified as, lateral lying, sternal lying, medium activity and high activity. Through the use of a rule based approach using a measure for activity taken from each axis on a per sample basis, accuracies of "up to 90%" were reported, although figures as low as 81% are also given for different classes. In [34], previous work was expanded and reported classification accuracies using a Multi-Process Kalman Filter described in [32] as between 75 and 100%. These results could be considered to be stronger than those presented in this study, however, it can be argued that the classes chosen for assessment in those studies are a key contributing factor. For instance Cornou et al. [34] aim to distinguish between high activity and medium activity which is a different problem to that addressed in this paper.

Escalante et al. [44] present an overview of a range of machine learning approaches to classifying sow behaviour, using the same data set described in [31, 34]. They describe a five class problem and aim to differentiate between Feeding, Rooting, Walking, Sternal Lying and Lateral Lying. The top performing classifier tested, logitboost, correctly classified an average of 74.64% 1-second observations. When considering 2-minute series, the logitboost classifier averaged 80% of series accurately classified. The primary source of confusion in the logitboost classifier was found to be between the lying categories, with good discrimination being achieved in the active categories. Again this is a slightly different task than the one presented in this paper.

Based on the range of movement available to sows housed in farrowing crates, the classes chosen for this work provide a detailed characterisation of the posture of the sow. Whilst altering these classes to simplify the problem could produce better classification performance it was felt that to be complete the system should classify to as detailed a level of posture as possible. Given that this is the case, however, specific drawbacks are encountered. Figure 3.4, for example, shows that the posture classification system developed in this work robustly identifies periods in which the sow is lying laterally. It can be seen, however, that classification between Sternal Lying and Standing as well as between Sternal Lying and Sitting is problematic. The class specific F_1 score for Sitting highlights this particularly (Table 3.2). Sitting can be seen to be a largely transitory state i.e., the sow sits for a short period between postures. This is similar in nature to the Kneeling posture, however due to the increased frequency and duration compared to Kneeling, it was decided to give Sitting a discrete classification. Despite this, the proportion of the data in which the sows are in the Sitting posture is substantially lower than the other postures, as such there is considerably less data with which to train the classifier for Sitting. Ringgenberg et al. [124] attempt to classify sitting behaviour and also found this to be a particularly difficult class to predict, correctly predicting 37% of sitting postures. In this work sensors were secured to a hind leg in addition to the sows' backs for a period of 6 hours. This approach produces very good results for classifying Standing (99.6%) and Sternal (93.5%) and Lateral Lying (96.7%), however, for longer term deployments this might be unsuitable as the sensors would likely be removed by the sows, as they describe. Cornou et al. [34] conflate Sitting, Standing and Sternal Lying into a "Medium activity" class, which again produces the least compelling results of the classes identified. This highlights the difficulties associate with correctly predicating this class, and identify an open problem for further investigation and improvement.

The transition frequency plots show the cumulative amount of transitions for each 2-hour period providing clear descriptions of the level of activity exhibited by the sow. As can be seen in Figure 3.6, the transition frequency plots from Pig106, Pig107, Pig235, and Pig252 clearly show the dramatic increase in activity as the sows attempt to exhibit nest building behaviour. Manual checks of the video recordings showed that these spikes in activity occur in the 12 to 18-hour period before the onset of farrowing. The peak rate of positional changes has been reported as being in the 6 hours prior to the onset of farrowing [91], although 12 hours prior to farrowing has also been suggested [146]. In other work in the field, activity data similar to this, has been used to predict the onset of farrowing [97, 111] and, given a larger dataset, it would certainly be possible to use the transition frequency metric generated in this study to predict farrowing. An accurate prediction of the onset of farrowing would provide farmers the ability to provide timely neonatal piglet care without reliance on continual monitoring. Detection of prolonged farrowing in particular can be used as an indication that human intervention is required [152, 157].

One of the biggest problems in the global swine industry is piglet mortality. This is greatest in the immediate post-partum period but may continue until weaning. Historically, it has been reduced by changing the environment of the sow by confining her to a crate, thus reducing her movements and providing safe areas for the piglets [35]. Even in these environments, however, sows exhibit nesting behaviour. Current research suggests that if nesting behaviour has not been carried out to the satisfaction of the sow, nest building may continue into parturition, increasing the amount of risk presented to the piglets [37]. Currently, lying quality and activity of sows is measured visually and temporally, through live observation or videos and recording the latency to complete each of the stages of lying [96]. In order to fully understand lying behaviour we need to be able to measure not only the time taken for a sow to lie but also the accelerations involved in the process. Through the use of our system it is possible to generate these data automatically, without the time investment required for manual observation.

Studies investigating the genetic basis of crushing have reported heritabilities for sows that are low, commonly ranging between 0.01-0.04 [48, 51, 52]. However, the risk of crushing seems to be constant over time for a given sow. Even with relatively low heritabilities, it would be possible to select for sows with lower levels of crushing [151] Baxter et al. [11], showed that a key factor towards distinguishing sow selection lines more prone to crushing appears to be care in movement. The method we present in this paper would be extremely useful in determining the degree to which the posture change traits associated with crushing in sows are exhibited and would consequently allow for selection of production sows based on these criteria.

Further development of the systems described in this work should consider the choice of features extracted from the transitions. Refinement, or expansion, of the features selected would allow for different analyses to be conducted. The features used in this study focus on describing the control, or lack thereof, exhibited in the sow's lying behaviour. Other analyses could be employed to assess the frequencies involved in transitions, the behaviour before and after the transitions, or the spatial context of the transitions. The algorithms developed for this study have been designed to be easily extensible should these analyses be required.

3.5 Conclusion

We have presented a framework for the automated detection and assessment of sow posture change, as well as an approach for sow posture classification through the analysis of inertial measurements. Through this we have provided access to data relating to the characteristics of sows lying behaviour during parturition. This kind of data could prove to be invaluable for several different applications, notably in characterising the nesting activity of sows' before and during farrowing, and in identifying sows with a predisposition towards posture changes that pose a danger to new-born piglets post-partum. The algorithms described herein have been designed with extensibility in mind from the outset. The implementation and addition of further behavioural features can be conducted and integrated simply and quickly, allowing for rapid assessment of lying behaviour with objectives different from those identified in this study.

4

Freedom To Lie

How farrowing environment affects sow lying behaviour assessment using inertial sensors

Contents

Abstract

We address the use of accelerometery to automatically monitor lying behaviour in free-farrowing sows; due to their freedom of movement and the consequent increased variety of movements the sows are able to exhibit, the challenges in automating this are greater than in sows housed in movement restricting farrowing environments. The methodology developed was applied to two salient applications: that of farrowing prediction through detection of nest building activity, and comparison of maternal lying behaviour in conventional movement-restricting and free-farrowing systems. Two sensors were attached at both the front and hind end to each of eight periparturient sows. Movement behaviour was recorded for a period of five days around parturition. Activity transitions were classified by a Support Vector Machine classifier, using data from both sensors individually, and combined; classifier output was validated against ground truth annotations collected from video data. We draw conclusions about the benefits of using multiple sensors over a single sensor, as well as the suitability of different sensor locations on the sow. Activity classification was found to improve through the use of multiple sensors, with a mean F_1 score (a measure of predictive performance between 0 and 1) of 0.84, compared to use of the front sensor alone (mean $F_1 = 0.49$) and the hind sensor alone (mean $F_1 = 0.57$). Activity transitions were classified using the dual sensor setup with a mean F_1 score of 0.77. Using a threshold-based approach, taking transition frequency as an indicator of nesting behaviour, we were able to detect the onset of nest building with an average latency to farrowing of 11.1 ($\pm$4.65) hours, and an average of 1 premature detection per sow; however, the majority of these premature were in a particular sow. We draw comparisons between the lying behaviour of free-farrowing and restricted sows. Using a mixed-design ANOVA we found a main effect of farrowing environment on transition duration ($p = 0.003$), peak acceleration ($p = 0.007$) and rate of change in pitch ($p = 0.009$). Improving the classification accuracy of sow activity transitions through the addition of multiple sensors allows for improved performance in applications such as farrowing prediction, which has the capacity to reduce piglet mortality through enabling farrowing supervision. Understanding how movement restriction affects the lying behaviour of farrowing sows

has the potential to inform decisions regarding restriction of sows and development of free-farrowing environments.

4.1 Introduction

The use of accelerometers to quantify animal behaviour has become widespread over recent years [67]. Studies have been conducted to investigate a range of animal behaviours, from routine activities such as running and playing in dogs [82], to more specific behaviours targeted at assessing welfare, such as lameness assessment in cows [116]. Due to their small size and the versatility of the data produced, accelerometers have been found to be particularly effective at monitoring animal behaviour in large-scale settings [100].

Automatic quantification of posture and lying behaviour has potential to enhance the welfare and productivity of various domesticated species. Changes in posture and activity may provide indications of underlying health and welfare issues [102, 135, 150]. Changes in activity and lying behaviour of pregnant sows can be indicative of the onset of farrowing, allowing for intervention and supervision, or to identify sows that pose less risk to their piglets by their lying behaviour [95, 130]. Accelerometry that quantifies the behaviour of sows during the farrowing process has used single sensors on animals [31, 111]. The use of multiple sensors to perform activity recognition in sows has been undertaken [124], in which one sensor was mounted to the back of the sow, as well as one secured to the rear leg of the animal. The leg worn sensor was targeted at assessing stepping behaviour, and postural assessments only utilised a single sensor at a time. Prior to the work described in this paper, we also conducted experiments into the use of accelerometry to quantify maternal lying behaviour of sows housed in farrowing crates, using a single accelerometer secured to the hind end of the pig [136].

It is standard practice in pig systems to move a parturient sow to a farrowing crate several days prior to the expected date of farrowing. This practice improves the survival rates of the piglets [35], however has been shown to increase stress in the sow [87] and suppress natural maternal behaviour [37, 65]. Farrowing crates are designed to restrict the movement of the sow, and as such approaches to automatically classify and quantify sow behaviours may be relatively straightforward. This restricted repertoire of behaviours reduces opportunity for misclassification between behaviours and allows for a more simplistic approach to classification.

On the other hand, alternative free-farrowing systems, such as PigSAFE [43] aim to optimise welfare and economic performance by allowing increased freedom of movement and expression of natural behaviour, whilst providing enhanced safety features for new-born piglets. When allowed free movement, the problem of automated behaviour classification becomes considerably more complex. Additional behaviour states must be considered and can be expressed with fewer physical constraints, specifically behaviours that involve movement from one area of the pen to another. Understanding the differences in sow lying behaviour between movement restricting and free-farrowing environments has potential impact on the management of farrowing. Quantification of these effects, if possible, may also have implications for promoting uptake of higher welfare systems.

Throughout this paper, we will refer to the combination of posture state and moving behaviours as "activities", and the period in which a sow moves from one activity to another as "activity transitions". It is the aim of this work to explore the potential for increasing activity transition classification accuracy through the use of multiple accelerometers and quantify improvement in the two farrowing systems. The methodology was applied to two salient applications that of farrowing prediction and comparison of maternal lying behaviour in movement restricting and free-farrowing systems. This allows us to demonstrate the applicability of the approach in scenarios that have practical significance. We hypothesised that the improvement in accuracy through the use of two sensors will be mainly in the alternative farrowing system.

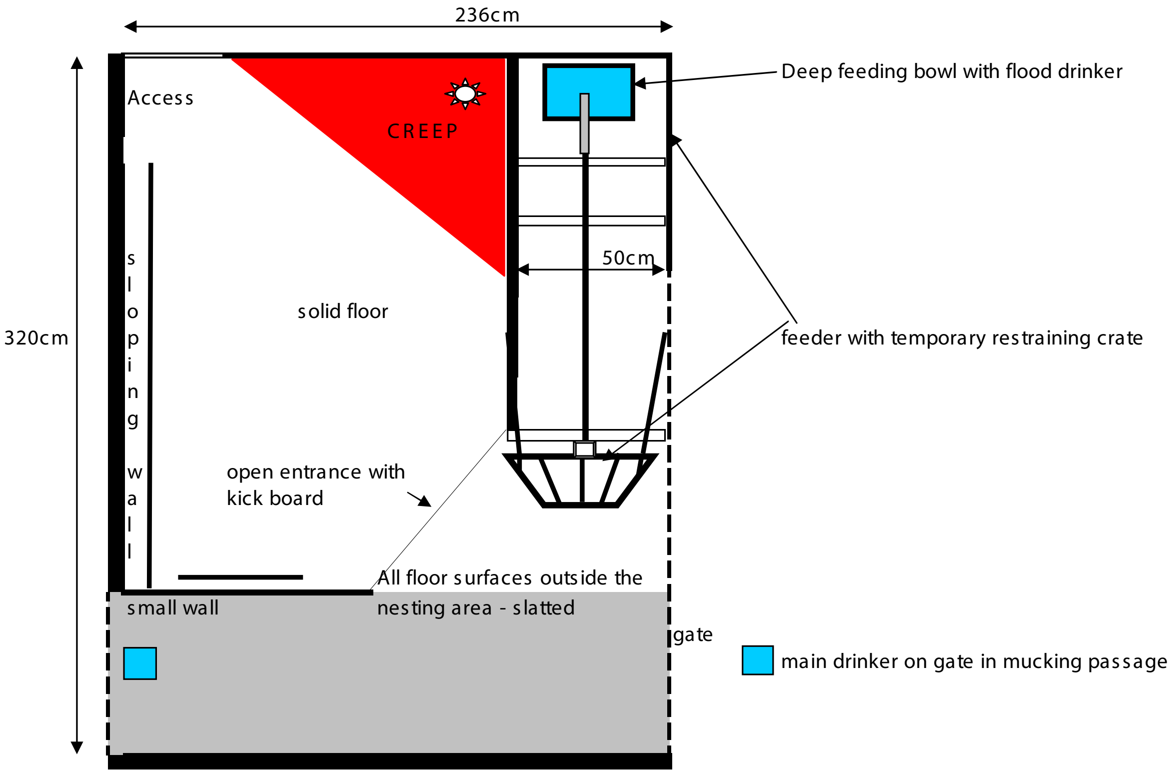


Figure 4.1: Floor plan of the free-farrowing (PigSAFE) pens. The sow is given room to move freely, as well as separate areas for dunging and feeding. Dedicated creep space is provided for the piglets and is fitted with a heat lamp [22]. Used with permission.

4.2 Materials and Methods

4.2.1 Data collection

4.2.1.1 Study animals

Eight hybrid sows from the same batch at Newcastle University Cockle Park pig unit were used for assessment and moved to farrowing accommodation three days prior to the expected date of farrowing. The sows were between 2nd and 4th parity. They were housed in PigSAFE farrowing environments[1], providing them with freedom of movement throughout the farrowing process. A floor plan for the PigSAFE systems is shown in Figure 4.1. Motion data was collected from the sows for five days during which the sows were housed in the PigSAFE system.

[1]https://www.freefarrowing.org/info/5/individual_farrowing_pens/1/pigsafe

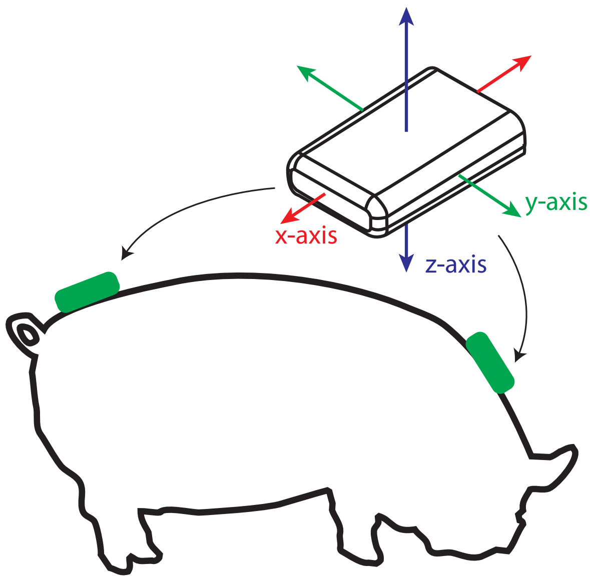


Figure 4.2: Position of the Axivity AX3 sensors used to monitor acceleration of the front and back ends of the sow. Both sensors were securely attached to the sows using a combination of adhesive tape and contact adhesive; they were aligned to the same orientation to ensure consistent measurements. Modified with permission from [136].

4.2.1.2 Sensor protocol

All procedures were approved by the Newcastle University Animal Welfare Ethics Review Board. Two Axivity AX3 triaxial accelerometers [8] were secured to the sows using a combination of contact adhesive, and adhesive tape. Sensors were secured to the sows during their morning feeding and prior to moving to the farrowing environment. All sows were shaved and cleaned in two locations: at the rear, just above the tail head and between the hip bones; and at the neck, halfway between the shoulder and the base of the skull, see Figure 4.2. Sensors were wrapped in duct tape to provide further protection should they become dislodged. A coating of Evo-Stik contact adhesive was applied in a $2cm$ patch around the sensor which was further secured with a layer of Scapa Sniper tape. Both sensors were aligned to ensure axes of measurement were consistent between sensors. The sensors were calibrated to collect data at $30Hz$.

Video data were recorded and annotated using the same protocol described in Section 3.2.1.2, deviating only in the addition of a second camera to ensure full coverage of the larger pen area afforded to the subjects. In addition to synchronising the sensor data with the video, the two video sources were synchronised together in a similar manner.

4.2.2 Lying behaviour assessment

Assessment of the sow lying behaviour was conducted on two levels:

1. Activity classification

2. Transition detection, segmentation and classification of activity transitions

Before this analysis could be conducted the data was cleaned and pre-processed to ensure that only data salient to the lying behaviour were included.

4.2.2.1 Data pre-processing

The data for each sow were then partitioned into short, overlapping frames of 1 seconds of data with an overlap of 0.8 seconds (24 samples). From these frames features describing the properties of the data were extracted. The features extracted were as follows (explanation below):

1. Mean pitch;

2. Mean roll;

3. Mean magnitude of acceleration;

4. Inverse Empirical Distribution Function (ECDF) coefficients;

5. Peak absolute acceleration.

The sensors record acceleration data in three spatial axes, hereon denoted as x, y, and z within the range $\pm 8g_0$ where $1g_0$ is the magnitude of the gravitational pull of the Earth at sea level ($-9.81ms^{-2}$). In order to calculate the pitch and roll of the sensors it was first required to estimate the static acceleration of the sensor, that is the acceleration caused solely by the gravity. Given a raw signal with each sample $\mathbf{s}$ at each time point (t) using a moving average filter, the static acceleration for each sample, $g(t)$, can be calculated as described in Equation 3.1.

where α is a weighting component used to vary the amount that the signal compensates for rapid changes in acceleration.

Subsequently, we were able to calculate the pitch and roll of the sensor. These measures describe the orientation of the sensor in 3-dimensional space relative to the vector describing acceleration due to gravity. Given the placement of the sensors as shown in Figure 4.2, pitch, p represents rotation around the mediolateral axis of the sow, and roll, r, represents rotation around the craniocaudal axis. Pitch was estimated using Equation 3.2, and Roll was estimated using Equation 3.3.

The magnitude, $m(t)$, of each point of the signal describes the overall amount of acceleration experienced by the sensor, independent of the direction of the acceleration and is calculated using Equation 3.4. The ECDF coefficients provide a representation of the distribution of the sensor readings, reducing its dimensionality whilst preserving the essential information. The method for calculating these coefficients is presented in [56].

The peak absolute acceleration describes the largest acceleration, either positive or negative, to occur in the signal, and was calculated as:

$$max(x(t), y(t), z(t)) \tag{4.1}$$

for all t within the frame.

4.2.2.2 Activity classification

Activity classification was performed through the use of a Support Vector Machine (SVM) classifier using a Radial Basis Function (RBF) kernel, trained on features extracted from the aforementioned 1-second frames of data. Parameters of the SVM were optimised using standard sequential minimal optimisation and hyper-parameters were optimised using grid-search [153]. The classifier was trained and evaluated using a leave-one-subject-out approach, in which the classifier was trained on data from all but one of the sows and tested on the other, see Section 3.2.6.2 for more details. The classifier predicted activity for each of the 1-second frames based on the feature vectors described in the previous subsection. The activities classified were: Left lateral lie (LL), Right lateral lie (RL), Sternal lie (SL), Sitting (S), Standing (ST), Walking (W), and Feeding (F).

A comparison between the use of a single sensor and two sensors was conducted by training optimal classifiers using: i) Front and rear sensor data, ii) front sensor only, and iii) rear sensor only.

4.2.2.3 Transition detection, segmentation and classification

Following activity classification, transitions between activities were identified. Initially the vector containing activity classification labels was smoothed to remove short period misclassifications. This process involved identification of consecutive data frames in which activity classification is inconsistent. Given the short nature of some transitions three consecutive frames were assessed in order to avoid smoothing over short transitions. Where any three consecutive frames were not found to have the same classification, but the initial and final frame of the three were the same, the middle frame was reclassified to match the outer two. Where all three frames were found to be different the subsequent frame was added to the assessment; if this was found to match the final frame of the previous three, the frame subsequent to that was added and if again found to match, the section is marked as containing a transition (see Figure 4.3 for a visual explanation of this process). Once stable activity classification had been achieved, the start and end of the transition were marked. The transition was then classified as being between the two stable activities either side of the transition.

4.2.3 Transition analysis

4.2.3.1 Kinematic characteristics of activity transitions

To perform analysis and comparison of lying behaviour between sows, kinematic characteristics of the transitions were extracted from the segmented transitions. These features included:

1. Duration of transition between activities - This was calculated from the end of the final frame of the previous posture to the start of the proceeding stable activity.

2. Rate of change of pitch and roll in both sensors separately - The pitch and roll of the accelerometer data describing the transition were calculated as described

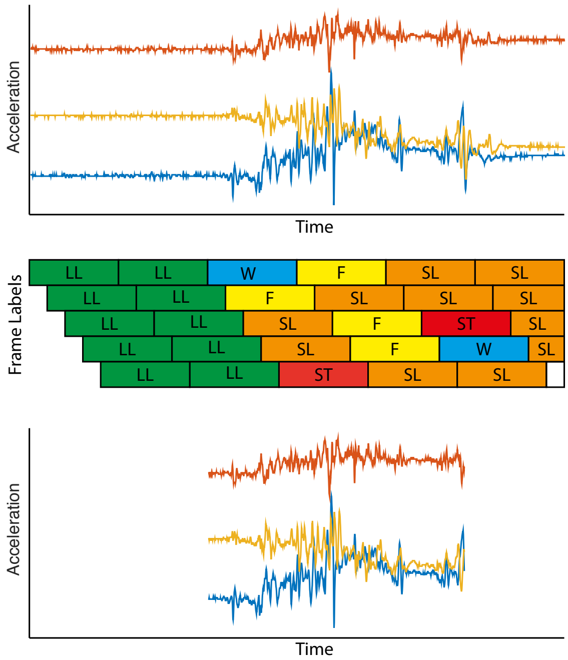


Figure 4.3: Transition segmentation and classification process. Top - raw acceleration data describing a transition from one activity to another (Left Lie to Sternal Lie). Middle - a representation of the frame by frame labelling of the data according to activity; each box represents a frame of data, coloured and labelled according to activity as predicted by the classifier. Labels included in this figure are LL: Left Lateral Lie, SL: Sternal Lie, W: Walking, F: Feeding, ST: Standing. Frames are staggered due to overlap of data within each frame. Bottom - The final segmented transition used for further analysis.

in Section 4.2.2.1. The rate at which these values change through the duration of the transition is used to describe the smoothness of the activity transition. These values are calculated as the first derivative of the pitch and roll for both sensors separately.

3. Rate of change of acceleration in both sensors separately (Jerk) - The rate of change of acceleration through the duration of the transition is calculated. This describes the smoothness with which the sow starts moving or comes to a halt. This is calculated for both sensors.

4. Peak acceleration and deceleration of both sensors separately - The maximum and minimum values recorded by the sensor across all three axes. This is calculated for both sensors.

5. Range of acceleration of both sensors separately - The difference between the

maximum and minimum values across all three axes.

6. Peak signal magnitude for both sensors combined - The maximum value of the magnitude of the signal when both sensors are combined. Magnitude is calculated as described in Equation 3.4.

7. Peak difference between pitch and roll between sensors - The difference between the maximum absolute pitch and roll between both sensors. This describes the difference in the orientation of the two sensors, and consequently the front and back end of the sow.

4.2.3.2 Comparison between restricted and free-farrowing sows

An *ex post* comparison was performed between sows housed in movement restricting systems and those housed in the PigSAFE pens. The data set describing the lying behaviour of those sows confined to movement restricting systems was collected for a previous study [136]. The sows in the restricted system study were six hybrid sows of either 2nd and 3rd parity, from the same stock of Newcastle Cockle Park pig unit used in the PigSAFE system. The sows were confined to farrowing crates for the period around the expected farrowing date and were fitted with a single sensor on the rear of the sow. Transitions were segmented, and kinematic features were extracted. The data collection protocol was essentially the same as in the current study, with the exception of the attachment of the second sensor.

To compare the transition characteristics of both sets of sows, a slightly modified feature set must be used from that described above. Features 1-5, taken from the rear sensor are common to both sets of sows and are therefore suitable for comparison. Similarly, due to the restricted movement imposed on the sows confined to farrowing crates, certain transitions are not exhibited by the sows. Consequently, only transitions between the postural states exhibited by both restricted and free-farrowing sows were considered for analysis, these were standing, sitting, lateral lying on left and right, and sternal lying.

Due to the variance between the size of the data sets, a random subsample of transitions was extracted for use in the analysis, to provide a balanced class distribution between data sets. Comparison of the transition profiles of both sets of sows was performed using mixed-design ANOVA to determine the significance between groups on each of the transition features, with a between-group factor of farrowing environment and repeated measures of transition within sow. An α level of 0.05 was used to indicate significance.

4.2.3.3 Prediction of the onset of farrowing

Increased transition frequency was used to detect nest building behaviour as a predictor for the onset of farrowing [35]. Transition occurrence was measured across the full duration of the study. Transition frequency was calculated using a moving average of transition count per 3-hour period in increments of 12 minutes. [5] found that false warning periods on average lasted 0.7 hours, this two-hour period should eliminate false warnings due to normal exploratory behaviour. Baseline transition frequency was estimated per pig by calculating the median 3-hour transition count for the 24-hour period between 36 and 12 hours prior to the delivery of the first piglet. An increase in transition frequency of more than twice the baseline level for ten consecutive 3-hour timeframes was used to indicate nest building and the impending onset of farrowing. Due to the 12-minute increments this constitutes a period of 2-hour increased transition frequency. Further detections were not flagged until the transition frequency had dropped below this threshold for ten consecutive 3-hour windows.

4.2.4 Evaluation and validation

To evaluate the accuracy of the classifier trained on activity data for the sows, a leave-one-out approach was taken, in which the classifier was trained using labelled data from seven sows and tested on the data of the eighth. The mean of the evaluation measures across all eight experiments was taken and is reported, giving an indication of the classifiers ability to accurately predict sow activity given previously unseen data. Three standard metrics for classification accuracy were produced: precision, recall, and F_1 score, see Section 3.2.6 for a description of how these are calculated.

Validation of the prediction of the onset of farrowing was performed retrospectively by reporting the mean difference between the last time activity passed the predetermined threshold before farrowing commenced and actual time of farrowing. False predictions were defined as any detection of nest building activity other than the detection that occurs closest to the onset of farrowing. False detection rates were reported to provide an indication of when the system might fail.

4.3 Results

4.3.1 Lying behaviour assessment

4.3.1.1 Activity classification

Given 5 frames of data per second for 6 hours per day, for all 8 pigs on each of 5 days there were a total of $4,320,000$ frames of activity data to be classified. Using a leave-one-pig-out cross validation technique the F_1 scores for the three sensor setups are described in Table 4.1. For all behaviours F_1 scores were higher when two sensors were used as opposed to one. Mean F_1 score for the combined front and rear sensor setup was 0.84, whilst mean F_1 for the front and rear sensors alone were 0.49 and 0.57 respectively. Classifications from the dual sensor setup were high across the board, with the exception of Feeding. The front sensor setup produced comparably higher F_1 scores for the lying activities, however there was considerable misclassification between Standing, Sitting, Walking and Feeding activities in which the sows exhibited more independent movement of the head. The rear sensor setup produced high F_1 scores for the lateral lies; however, it produced lower results for Standing, Sternal Lie and Walking, in which the sensor is in essentially the same orientation for all activities. Analysis of the rear sensor data also produced significant confusion between Feeding and Standing.

4.3.2 Transition classification

In the 240 hours of annotated video there was a total of 481 transitions. Having identified the dual sensor setup as the most accurate based on the activity classification results (see Section 4.3.1.1), the generated postural labels were used to perform the transition classification. An overview of the transitions present in the data set can be

Table 4.1: Mean F_1 scores (a measure of predictive performance for classification accuracy, ranging between 0 and 1) for each activity class under different sensor setups for 8 sows housed in a free-farrowing environment. Classification was performed on 6 hours of data from each of 5 days, totalling 240 hours.

Sensors	Standing	Sitting	Left Lateral Lie	Right Lateral Lie	Sternal Lie	Walking	Feeding	Mean
Front & Rear	0.77	0.96	0.98	0.97	0.78	0.84	0.60	0.84
Front Only	0.41	0.35	0.74	0.77	0.39	0.31	0.48	0.49
Rear Only	0.57	0.42	0.86	0.91	0.48	0.52	0.26	0.57

seen in 4.2. The mean F_1 score for transition classification was 0.77. Transition classification suffered where the postural classification that was used had lower accuracy, as can be seen in transitions involving standing activities. Transitions to a right lateral lie were also notably low.

Table 4.2: Mean F_1 scores (a measure of predictive performance for classification accuracy, ranging between 0 and 1) for transition classification for all transitions included in the dataset. Transitions that were not exhibited during the period annotated are marked with a dash. Activities transitioned from are listed vertically, activities transitioned to are listed horizontally.

From\To	Sit	Stand	Left Lie	Right Lie	Sternal Lie	Walk	Feed
Sit	-	0.81	0.92	0.45	0.98	0.66	-
Stand	0.71	-	0.6	0.54	0.84	0.78	0.88
Left Lie	0.87	0.62	-	0.52	0.9	-	-
Right Lie	0.96	0.66	0.46	-	0.93	-	-
Sternal Lie	0.72	0.78	0.86	0.92	-	0.85	-
Walk	0.58	0.72	-	-	0.59	-	0.96
Feed	-	0.92	-	-	-	0.98	-

4.3.3 Comparison of lying behaviour between movement restricted and free-farrowing sows

The output of the mixed-design ANOVA to quantify the difference in transition features between movement restricted and free-farrowing sows are shown in Table 4.3.

The ANOVA revealed a main effect of farrowing environment on transition duration ($p = 0.003$), peak acceleration ($p = 0.007$), and rate of change in pitch ($p = 0.009$), whereas Range of Acceleration, Jerk and Rate of change of roll showed no significant

Table 4.3: Output of Mixed-design ANOVA for transition features between restricted and free farrowing (PigSAFE) housed sows during the period around farrowing. Mean square errors, the F test statistic with 1 degree of freedom between group and 12 degrees of freedom within group and significance values are reported. Effects that exhibit statistical significance at the $\alpha = 0.05$ level are marked with an asterisk.

Transition Feature	Mean restricted	Mean PigSAFE	F(1,12)	p
Duration	11.99	9.62	13.52	0.003*
Maximum Acceleration	0.20	0.27	10.47	0.007*
Range of Acceleration	2.87	2.76	0.435	0.522
Jerk	0.00018	0.00024	1.051	0.326
Rate of Change of Roll	0.037	0.202	2.051	0.178
Rate of Change of Pitch	0.039	0.331	9.523	0.009*

effect of farrowing environment. Mean values for each of the transition features are also shown in Table 4.3, and show that restricted sows exhibited significantly longer transition durations, lower peak accelerations and decreased rate of change of pitch during transitions.

4.3.4 Farrowing prediction through nest building activity

Nest building detection as an indicator of start of farrowing was performed on the 8 sows in the free-farrowing environments, by identification of increased activity transitions which had been shown to precede farrowing in a previous study [136]. Nest building detections were marked where activity transition frequency exceeded a threshold based on baseline data. The mean time between the final threshold passing event and the onset of farrowing was 11.1 ± 4.65 hours. Using this method, the algorithm predicted the onset of farrowing within 8 hours for 4 of the sows - sows 2, 4, 6 and 8. Nest building was detected in the other four sows between 10 and 17 hours before the actual farrowing began. The algorithm incorrectly detected nest building activity for 5 sows, registering increased transition frequency above the threshold, more than once before farrowing commenced. Detections were determined to be incorrect if registered prior to a subsequent detection and before the onset of farrowing. Data from 4 sows produced only a single activity transition frequency threshold passing event prior to the onset of farrowing. A visualisation of the nest building detection can be seen in Figure 4.4. Farrowing was predicted prematurely in several cases due to increased activity immediately following rehousing (Sows 6, 7 and 8).

4.4 Discussion

Whilst alternative farrowing accommodation is currently only employed for a relatively small proportion of farrowing sows within the pig industry, this is set to increase due to increasing concerns over sow welfare [43]. The advantages of being able to quantify sow lying behaviour have been explored in movement-restricted sows in a previous work [136]. There we showed that lying behaviour assessment has the potential to identify sows with a predisposition to posture changes that present a risk to new-born piglets and may be able to provide early warnings of the onset of farrowing. In the current work, we set out to explore the challenges presented by activity recognition and activity transition assessment in free-farrowing sows. We hypothesised that, due to the increased range of movement, the activity and transition classification challenge would be more complex, and as a result accuracy would suffer. We identified the use of an additional sensor to the sow as a potential approach to improve the reliability of the predictions.

In our previous work [136], we employed a single sensor to record sow movement. The motivation for the use of a single sensor at the hind end of the sow was to measure acceleration originating from the point at which the largest forces would be brought to bear, when the sow changed postural states, in particular when the sow performed posture changes considered to be more dangerous to the new born piglets [14]. Whilst this was found to be appropriate in monitoring behaviour, certain postures suffered from poor classification accuracy arising from the single point of measurement, namely sitting, sternal lying and standing. In this study, we have not only addressed this, but have also been able to quantify the improvement. We performed three experiments in which we drew comparisons between the number of sensors used and the location at which they were affixed to the sow. Using a single sensor on the hind end of restricted sows, posture classification algorithms produced a mean F_1 score across five postures of 0.78 [136]. It should be noted that the results presented in Chapter 3 and 4 are not directly comparable, as the classifiers are trained on different data, however the

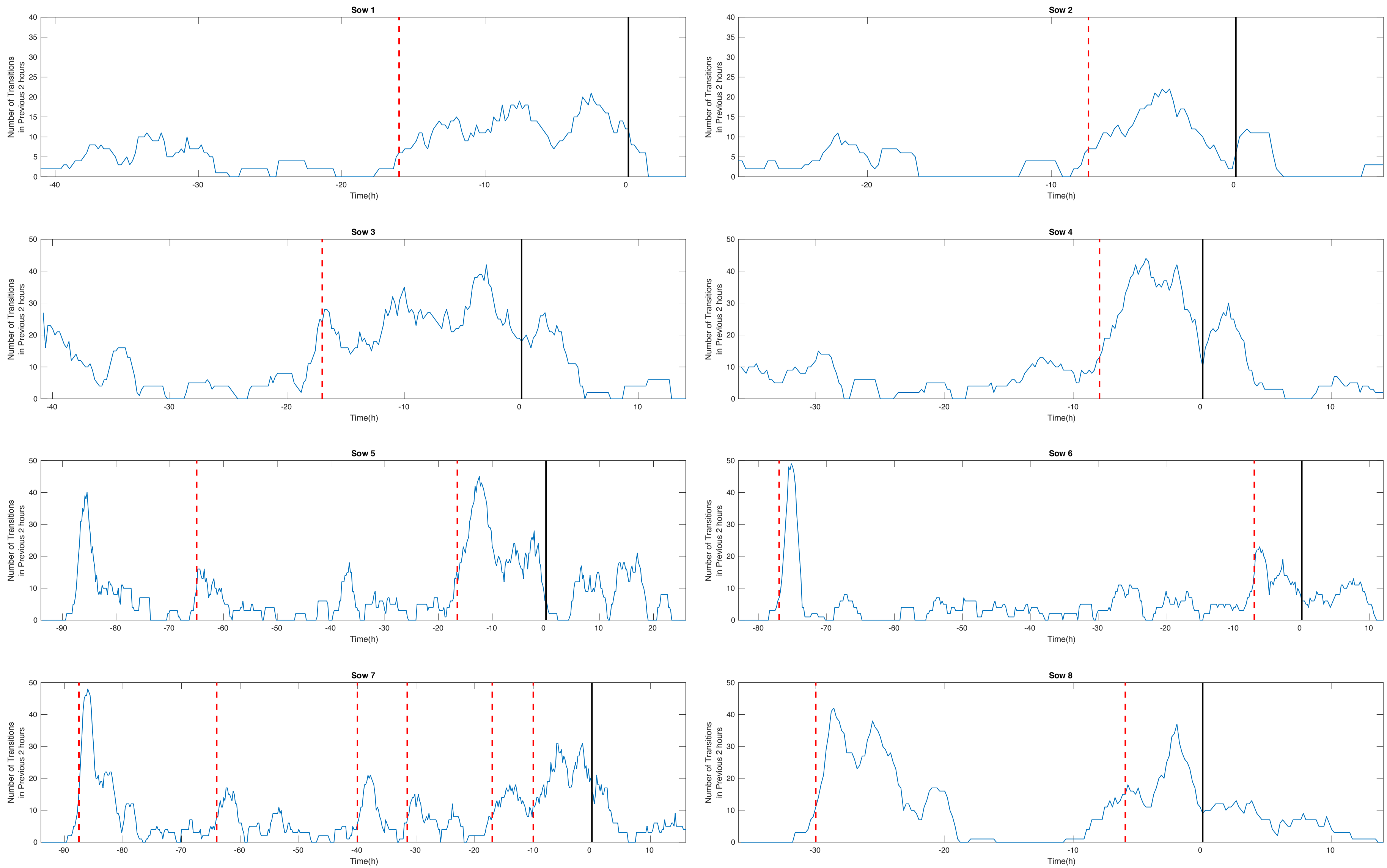


Figure 4.4: Transition Frequencies for all sows around the period of farrowing, including false alarms. The solid black line represents the point at which the first piglet is born. Dashed red lines represent detections of nest building behaviour indicating the onset of farrowing based on increased transition frequency (twice the activity transition frequency of the median for the 24-hour period between 36 and 12 hours prior to the birth of the first piglet in the PigSAFE environment). Charts are scaled on the horizontal axis to include the periods containing farrowing and threshold passing events.

performance metrics generated provide an indication of the potential for comparison between sensor numbers and positioning. In the present study, we conducted the same assessment on free-farrowing sows, now with a total of seven activities under assessment (the five original postures with the addition of feeding and walking), which is reflected in the algorithm mean F_1 score of only 0.57. Performing the classification on data recorded from a single sensor affixed to the front end of the sow, we achieved an even lower mean F_1 score of 0.49. This reduction in accuracy is not surprising given the similarity of the postures when considered from the perspective of the front end of the sow. However, when using a front-end sensor in addition to a rear-end sensor a much higher F_1 score of 0.84 was achieved, an outcome which was consistent with our hypothesis.

We set out to implement the activity and transition classification algorithms in movement restricted and free-housed sows in order to ascertain the improved accuracy that was achieved. We aimed specifically to draw a comparison between the lying behaviour of restricted and of free-farrowing sows. Extensive research has been conducted into the effect of the farrowing environment on maternal behaviour, using traditional methods of assessment, such as manual observation [11, 35, 58, 125]. More recently, work has been conducted in an attempt to quantify these differences. [114] aimed to predict farrowing in both restricted and free-farrowing sows using collar worn accelerometers, and found that the overall activity level of restricted sows is lower than that of free housed sows. The authors were unable to determine what the cause for this difference was, however, it was suggested that ability to exercise nesting behaviours could be responsible. We have assessed the difference between transition features between movement restricted and free-farrowing sows. Whilst several features showed no significant difference between the systems, we found that in the restricted sows transition duration was longer, whilst peak acceleration and rate of change of pitch were lower. These three factors describe a more controlled transition profile, which is to be expected, as sow speed of movement is restricted by the confines of the farrowing environment. This indicates that restricting sow movement is an appropriate way to ensure more controlled and potentially less dangerous lying behaviour. Whilst this is important for the reduction of piglet mortality, it comes at the cost of the sow ability to freely express natural behaviour, including pre-lying behaviour which has been shown to reduce crushing behaviours as the sow gathers the litter away from where she intends to lie [36]. By using the methodology described above to identify sows whose lying behaviour was not significantly affected in this way by movement restriction, it could be possible to highlight them as suitable for free-farrowing. This would however require a significant amount of research into how we define "good" lying behaviour.

We also aimed to apply the dual sensor algorithm to farrowing prediction in free farrowing sows. Generally, the onset of farrowing is predicted by identifying increases in activity, assumed to be caused by the expression of nesting behaviour [33, 111, 113, 115]. We have employed a similar technique in this work, however rather than generic activity, we considered postural changes as an indicator that commencement of farrowing is imminent [136]. The implementation of any farrowing prediction system must be able to run in real time and give an actionable call to intervention, if it is to be of practical use. Current efforts towards farrowing prediction still have a long lead time to farrowing, with peak activities being found between 4 and 8 hours prior to farrowing [115] and nest building activity beginning several hours prior to this peak. Similarly, in this work we have found the increase in transition frequency to commence between 3 and 17 hours prior to farrowing. Whilst this allows for the raising of alarms, the delay until farrowing is still too long to be of specific use as a method for prompting intervention. The process is used also prone to false detections. Particular bouts of high activity were misclassified as commencement of nest-building activity. The choice of baseline activity levels used to calculate the threshold by which we mark a nesting behaviour point has impact on the success of the detection algorithm. Initially we used the first 24 hours during which the sow was housed in the PigSAFE pens, however, we found this created elevated baselines due to the initial exploratory behaviour displayed by the sows in response to the new environment. By altering this baseline period to cover a more stable period (12-36 hours prior to the birth of the first piglet), in conjunction with increasing the duration of sustained activity require to trigger a detection, we reduced the number of false alarms. [111] used the increase in activity to trigger alarms every 2 hours, thus introducing an expectation of false alarms, until activity falls below a threshold, allowing for continuous updating on the status of the farrowing process. Using a combination of video recorded activity levels and water intake [5] we were able to provide a coherent warning that farrowing was due to commence within 12 hours in 97% of their test cases. It was found that mis-detections lasted on average 0.7 hours. Given the nature of sow behaviour prior to farrowing, improvements in time to farrowing prediction will likely require a significantly different approach. The increased freedom of movement afforded to sows through the use of alternative free-farrowing environments like PigSAFE also introduces a degree of uncertainty in the interpretation of sow pre-farrowing behaviour. As the sow is more able to exhibit rooting and exploratory behaviours, there is the potential for premature prediction of farrowing when using an activity-based method. This has been shown to be the case in this study and in the literature. Additional input of automatically generated biometric data, such as body temperature and heart rate could almost certainly aid in reducing this. The introduction of additional sensors, whether an extra accelerometer, as in this study, or biometric sensor, is bound to reduce the feasibility of practical applications. In an exploratory capacity, this should not act as a deterrent to experimentation however. Certainly prior to large-scale, on-farm deployments considerations of cost and effort required to install systems such as this, and those described in the literature, must be made.

4.5 Conclusion

We present a novel approach to classification of activity and activity transitions in farrowing sows, and assess the efficacy of various sensor placements and numbers. We found that through the use of multiple sensors we were able to achieve higher levels of accuracy for activity classification, and that if only a single sensor is to be used, it is better placed at the rear of the sow, rather than at the head end. We drew comparisons between movement-restricted and free-housed sows, demonstrating that movement-restricted sows exhibit signs of increased control during activity transitions, which could be a factor contributing to the reduced mortality of piglets to sows housed in movement-restricting environments. We implemented an approach to farrowing prediction based on transition frequency indicative of nest building and are able to predict the onset of farrowing with an average accuracy of approximately 10 hours. The methods presented in this paper could be exploited to assess the maternal behaviour of a much larger number of sows and could provide an approach to refining the selection of sows to be housed in either free-farrowing pens or farrowing crates.

5

A Cat's Life

Automated quantification of activity and behaviour recognition in the domestic cat

Contents

Abstract

We examine the use of accelerometers for the assessment of feline behaviour. Feline behaviour monitoring is a key aspect of veterinary treatment, leading to improved health and welfare outcomes. Continuous monitoring of an individual over a long period of time may not possible, due to the expense and time input required by the observer. As such we propose a system which is able to automatically monitor feline activity and classify the behaviours exhibited during observation. In the development of this system two datasets were collected. The first included 52 cats residing at the Hill's Pet Nutrition Centre in Kansas, USA. Data were recorded from a single collar mounted accelerometer for a period of 2 weeks. An analysis of the activity of the subjects was performed to identify patterns and differentiate groups of cats. The approach was validated through comparison with subjective assessments conducted through questionnaires completed by the facility staff. We found significant correlations between the age of the subject and its activity level. Activity profiles for the cats were generated and cyclic rhythms in the daily activity of the cats were observed. The second phase of the study collected accelerometer data from 31 pet cats under the supervision of their owners for approximately 15 minutes. The subject was encouraged to perform behaviours determined to be of interest to both the owner and veterinarians, such as jumping, scratching and elimination behaviour. Behaviours were classified using a Support Vector Machine classifier. Predictions from the classifier were validated using video data recorded by the owner during the data collection period. The classifier produced a mean F_1 score of 0.729 across all behaviours. The classifier performed best on behaviours involving movement of the subject (walking, running, jumping), and less well on behaviours in which the subject was stationary (sitting, lying, defecating, urinating). The results produced are in line with results detailed in the contemporary literature and, further, assess behaviours that have not been treated in this manner previously. Using this methodology we have identified applications for the technology for veterinarians looking to supplement their diagnostic tools, and owners keen to gain a better understanding of their pet's lifestyle.

5.1 Introduction

Monitoring of animal behaviour is an important part of maintaining wellbeing, aiding in diagnosis of challenges to health, and understanding the animal's disposition [50, 103]. Traditionally this is undertaken at specific times, through a planned visit to a veterinarian. This occurs, however, only when there is a noticeable issue that has prompted concern in the owner, as part of a clinical examination. Many feline diseases initially present with subclinical symptoms [16, 90, 145], early diagnosis of which would allow for much improved outcomes [117]. Due to the nature of subclinical disease, it is challenging to determine when an animal requires attention. Automated monitoring systems may present an avenue to improve detection of subclinical illness, in addition to the detection of clinical symptoms that have not been recognised by the animal's owner.

Monitoring of behaviour during a short clinical appointment may be associated with several disadvantages, beyond the one associated with the shortness of time. Cats may be prone to disguising ailments when observed or show untypical behaviours in an unfamiliar environment, such as restless behaviour or inactivity (freezing); and intermittent conditions may not present during clinician observation, leading to missed symptoms, for example. As such there would be substantial benefit to the implementation of a system for round-the-clock monitoring of the animal that was not dependent on human attention, alleviating time and expense constraints.

Analysis of accelerometry data captured from wearable sensors for animal behaviour monitoring has become widespread. The increasing ease of use, and reduction of cost of sensor platforms such as Vetrax[1], Fitbark[2], Whistle[3] and others demonstrate the feasibility of such a system for dogs. Whilst there are currently few commercially available sensors that achieve the same functionality in cats, academic research has been conducted into such systems and approaches.

[149] presented an investigation into the use of a uniaxial accelerometer for the classification of a single cat's behaviour. Behaviours targeted for observation were broadly

[1]https://www.vetrax.com
[2]https://www.fitbark.com
[3]https://www.whistle.com

categorised into grooming, eating and travelling. Whilst the accuracy of the classification system presented a solid proof of principle, the development of the algorithm using data from a single cat necessitates modification if it were to be applied generally, as differences between individuals would confound the classifier training. Work has also been conducted on the application of activity trackers to specific health conditions, notably osteoarthritis, musculoskeletal pain and other degenerative joint diseases in cats [13, 54, 55, 86]. The research in this field has relied, primarily, on the use of activity monitors, through which activity is characterised by a count of events per epoch. This is valuable information and provides insight into the cats' general activity level which have been shown to correlate with both pain indices and have been used to validate the performance of medications as a result.

A more general observational study of feline behaviour has been conducted in which a variety of sensing modalities were employed to monitor the behaviour of a single cat [156]. Yonezawa et al. used accelerometry, video, audio and GPS data in an attempt to characterise the senses (sight, hearing, smell) of the feline. Whilst their research does not present concrete implementations of all the topic studies, it presents an interesting overview of how various modalities can be combined to generate an overarching view of the cat's behaviour and interactions.

The work presented in this paper develops upon related work in which activity monitors have been used and extends it by developing algorithms for the identification of specific behaviours in the domestic cat. It is key that the tools created must be applicable across individuals and, as such, a large dataset containing data from tens of cats has been collected for use. The primary aim of the study is to automatically classify specific behaviours and identify behaviour trends in domestic cats. This aim can be broken down into three key objectives: i) to develop and validate an algorithm to measure activity trends in group housed cats; ii) to develop an algorithm for the classification of behaviours deemed to be important to the characterisation of the health and wellbeing of domestic cats, and iii) to identify challenges presented by the use of accelerometry in order to provide a basis for future multi-modal systems.

5.2 Methods

5.2.1 Study 1: Activity assessment of colony housed cats

5.2.1.1 Data collection

Two datasets were collected for use in this study. For the first of these, data were recorded using collar-mounted triaxial accelerometers recording at $100Hz$ with a sensitivity of $\pm 8g_0$. Sixty cats were included in this study, housed in mixed sex groups of between 10 and 15 individuals at the Hills PNC. Complete data were retrieved from 52 cats. Three of the initial 60 were excluded due to improperly set up sensors and 2 were excluded from the study due to faulty sensors and a final 2 were excluded from the study for unrelated causes. Of the complete data there were 31 females and 21 males. The age of the subjects ranged from 6 months to 17 years. The mean age of the cats was 6.36($\pm$5.31) years. Thirteen cats in the study had previously been diagnosed with some form of ongoing condition. The cat handlers were requested to complete subjective personality profiles for each animal. Data were recorded from the cats over a period of two non-consecutive weeks.

5.2.1.2 Sensor attachment protocol

The sensors were fastened to the collars first by orienting the sensor on the collar and wrapping tightly with cohesive bandage (see fig. 5.1 left). Sensors were subsequently secured to the cat using quick snap collars in the normal manner (see fig. 5.1 right). Collars were fastened securely however effort was made to ensure that they caused minimal discomfort to the animals. It was anticipated that the collars would cause some behavioural response in the cats, however it was decided that approaches to managing this would be incorporated into the data-pre-processing stage of algorithm development.

5.2.1.3 Data preprocessing

In order to allow a more accurate analysis of the data with regards to different behaviours and activity states, it was important to first ensure that data was clean

Figure 5.1: Left: AX3 sensor correctly orientated on feline collar (left bottom), and sensor labelled and secured to collar using cohesive bandage (left top). Right: Colony cat wearing sensor fitted collar.

and uncontaminated by non-relevant data. Firstly, we removed data that had been recorded but did not describe the movements of the cats, that is, data recorded when the sensor was not secured to the animal. Secondly, we needed to identify data that we determined to be misleading in terms of the activity level.

The first stage required that we identify the periods in which the collar had not yet been attached to the cat, and the periods in which the cat had removed the collar. The second stage refers to the identification of activities that may appear to involve a high level of activity whilst not actually being the result of exertion by the cat. This generally took the form of scratching events causing large acceleration readings from the sensor. Whilst we were interested in this behaviour in general, it was considered important to prevent these events contaminating the data for the purpose of activity profiling, as it was determined through reports from the colony technicians that this was primarily a result of scratching due to a lack of habituation to the collar.

5.2.1.4 Calculation of activity

The activity calculation was based on the determination of specific thresholds relating to the various activity levels displayed by the cats. The thresholds were applied to the standard deviation of the magnitude of the acceleration data across a 10-sample frame, representing 0.01 seconds of real time. The magnitude of the signal is calculated for the frame as described in Equation 3.4. Subsequently we calculated the activity value A as the standard deviation of M for the 10-sample frame.

We performed this calculation for the entire dataset, generating activity values for each 10-sample frame. We used this data to inform the thresholds we set for four activity levels: No activity, low activity, medium activity and high activity.

This approach provided us with a metric describing the activity level of the cat ten times a second. This level of resolution gave us a clear and accurate picture of the behaviour of the cat. To provide a value more easily visualised, allowing for simpler intuition regarding the underlying meaning of these values, we aggregated them in two ways.

Firstly, we assigned a weighting to each of the activity levels. Through analysis of the distribution of activity level data we observed that the quantity of values representing higher levels of activity are much less frequent that those at lower levels. As such we assigned them a higher weighting in order to ensure they are better represented in the analysis. The initial classification of high, medium and low activity labels was selected based on a quartile analysis of the acceleration signals across the entire dataset; however the weights were chosen intuitively, leaving the potential for fine tuning to ensure more accurate representation. This selection was informed by discussion with a veterinarian who indicated that the general relationship between energy expenditure for different levels of activity could be used as a starting point. The technicians rated the cats on a discrete scale from 1 to 3, where 1 indicates the cat is generally active ($n = 11$), 2 indicates the cat is neither especially active or inactive ($n = 26$) and 3 indicates that the cat is generally inactive ($n = 15$). The 4 levels chosen align with the 3 levels described in the technicians reports, with the exclusion of the "No Activity" level all cats displayed at least some degree of activity.

Secondly, for the purposes of developing the feline activity profiles (see Section 5.2.1.8), we took the sum of these values across a 15-minute window. This provided us with a value that allowed us to interpret the changes in activity over the course of a longer time period without obfuscating these trends with the noisier signal produced when activity is viewed in 10-sample frames.

5.2.1.5 Sensor habituation

To determine if the cats had habituated to the sensor laden collars we quantified the number of cats that removed their collars during the first week and the number that removed during the second week, based on the reports of the colony technicians. This showed no significant difference; however, it is worth noting that between the two weeks only three cats removed their collars in both weeks. In the first week two cats removed their collars more than five times, in the second week both of these cats removed their collars only once. The majority of cats did not remove the collar at all (see Table 5.1).

Table 5.1: Summary of collar removals between weeks as reported by colony technicians. Total number of removals denotes each recorded instance of a collar being removed across all cats

Week	Number of cats that removed collars	Number of cats that did not remove collars	Total number of removals
1	11	41	22
2	9	43	17

5.2.1.6 Activity correlations

By taking the mean of the 15-minute activity values produced during the accumulation process described in the in the Calculation of Activity section we were able to describe the activity level of the cats across the entire study. Pearson's correlation was calculated between the age of the subjects and the aggregated activity data.

We considered the activity level of the cats based on a several different sub-setting approaches. A key amongst these was to compare the activity level generated by our metric with the subjective description of activity as assessed by the technicians tasked with the day-to-day care of the cats. An independent samples t-test was used to compare the activity level of cats as described by our metric and the activity level as described by the technicians.

5.2.1.7 Additional measurements

Body condition score (BCS) is a scale by which the body composition and body fat mass of a cat can be assessed [84]. This allows us to quantify the physical condition of the cat independently of its age, size and weight. The technicians were asked to complete a body condition score evaluation, using a five-point scale for each of the cats in the study[4]. The cats in this study fell into 4 of the 5 groups: Underweight ($n = 1$); Ideal ($n = 17$); Overweight ($n = 16$) and Obese ($n = 1$).

BCS is not measured on very young cats as their bodies have not yet developed to the stage where it is considered meaningful. To assess the validity of this an independent sample t-test was performed to measure the difference between the activity of kittens ($n = 14$) and adults with an ideal BCS ($n = 17$).

5.2.1.8 Activity profiles

In order to gain insight into how the behaviour of the cats varies over the course of a day the median activity values for each 15-minute timeframe in a 24-hour period were calculated for all cats across all days and visualised to identify how the activity level of the entire population changes throughout the course of any given day. The data were smoothed to remove noise and patterns were identified. Activity plots were also generated to present an insight into the behaviour of individual cats within the cohort.

5.2.2 Study 2: Recognition of targeted behaviours

5.2.2.1 Data collection

The second period of data collection aimed to collect video data in conjunction with acceleration data to develop a targeted behaviour recognition system. Due to the more laborious data collection protocol, data were limited to short periods in which the cats were encouraged to exhibit a range of behaviours through the use of treats, food, and owner playing. These data were also recorded using triaxial accelerometers recording at 100Hz with a sensitivity of $\pm 8g_0$. Owners were requested to film their cat's

[4]https://www.aaha.org/public_documents/professional/guidelines/weightmgmt_bodyconditionscoring.pdf, accessed: 14/12/2018

behaviour over a period of approximately 15 minutes, whilst filming the activity using a smart phone (models varied). Sensors were secured to the cats using the protocol described in Study 1.

In order to determine which behaviours were of most interest a survey was conducted on 49 cat-owning employees of Hill Pet Nutrition Centre. The results of the survey concluded that behaviours of most interest to owners of indoor only cats were: Drinking, time spent in prohibited areas, eating, urinating and sleeping. Walking, running, scratching and defecation were determined to provide additional insight into the general wellbeing of the cat. We determined that classification of jumping to and from a height would be appropriate way to characterise time spent in prohibited areas as, particularly for indoor cats, higher levels such as countertops and furniture are often prohibited and the act of jumping up to a surface would, in some instances, indicate that the cat had entered a prohibited area.

Sensor data and video footage were synchronised using a standard procedure of clapping the sensor in front of the camera to produce a large clear acceleration signal that can be matched with the movement in the video [121]. Data were then annotated using the open source application ELAN [19]. Definitions of the behaviours annotated are provided in Table 5.2. Behaviours definitions provided to the annotator of the video data. Where the cat's behaviour was deemed to be ambiguous the annotator was instructed to use their best judgment or to label the activity as null.

5.2.2.2 Preprocessing and feature extraction

Accelerometer data were divided into windows of 0.8 seconds (80 samples) with a 0.2 second overlap (20 samples). This duration was chosen by identifying the shortest duration events that occurred in the data set and ensures that a short duration event will be fully contained within a single window, without using needlessly long windows. 10 ECDF features (Empirical Cumulative Distribution Function) were extracted from each axis in each window to form a feature vector to train the classifier. ECDF features have been shown to reduce the dimensionality of an accelerometry signal, whilst preserving the statistical characteristics [56]. To calculate the ECDF features the below process is employed.

Table 5.2: Behaviour definitions provided to the annotator of the video data. Where the cat's behaviour was deemed to be ambiguous the annotator was instructed to use their best judgment or to label the activity as null.

Behaviour	Definition
Jumping up	The cat moves vertically upwards to a level above its starting position in a leaping motion
Jumping down	The cat moves vertically downwards to a level below its starting position in a leaping motion
Running	The cat moves forwards at a running pace
Walking	The cat moves forwards at a walking pace
Playing	The cat is engaged in playful activity, either on its own or through interaction with an owner
Scratching	The cat scratches any part of its body using a hind leg
Eating	The cat consumes food from a bowl, this does not include eating treats delivered by the owner
Sitting	The cat remains in a constant sat posture for a period of more than 0.5 seconds
Lying	The cat remains in a constant lying posture for a period of more than 0.5 seconds
Defecation	The cat excretes faeces, either in a litter tray or elsewhere
Urination	The cat excretes urine, either in a litter tray or elsewhere
Unspecified	The cat is exhibiting a behaviour not listed, or is transitioning between behaviours for a period exceeding 0.5 seconds

Initially the empirical cumulative distribution function, P_c, is calculated:

$$P_c(x) = P(X \leq x) \tag{5.1}$$

We then select d equally spaced points. For each of these points $p_i \in \mathbb{R}_{[0,1]}$ between 0 and 1, we estimate the value x_i for which each $P_i(x_i) = p_i$. For a collection of points C belonging to window i, the ECDF feature vector f_i becomes:

$$C = \{p_i\} \in \mathbb{R}^d_{[0,1]}, \quad p_i < p_{i+1} \tag{5.2}$$

$$f_i = \left\{x, \exists j \colon P^i_c(x) = p_j\right\} \tag{5.3}$$

In the current application d is set to 10, and the ECDF features are calculated for each axis of the acceleration signal giving us a 30-dimensional feature vector for each

240 sample (80 samples in 3 dimensions) analysis window.

5.2.2.3 Behaviour classification

In order to classify the feline behaviours of interest automatically, a Support Vector Machine (SVM) classifier using a Radial Basis Function kernel was trained on the feature vectors extracted from the windows of accelerometry [153]. Parameters were optimised using standard sequential minimal optimisation and hyper-parameters were optimised using grid search. To evaluate the performance of the classifier a leave-one-subject-out approach was taken, in which the classifier was trained with data from all but one cat, and subsequently tested on the unseen data. This is repeated, testing on data from each cat in turn. Mean values for precision, recall, and F_1 score are reported to provide an indication of the classifiers ability to determine behaviour from unseen data, see Section 3.2.6 for descriptions of how these are calculated.

5.3 Results

5.3.1 Study 1: Activity assessment

5.3.1.1 Validation against technician observations

A significant difference was found between the activity levels of the generally active group and the other two groups ($p < 0.001$) as classified by the technicians; however there was no statistical difference between the group that was neither active or inactive and the group that was generally inactive ($p = 0.147$). This suggests that our metric agrees with the subjective reports of the technicians with regards to the highly active cats (see fig. 5.2).

5.3.1.2 Activity correlation with age

The correlation between the age of the subjects and the 15-minute aggregated measure of activity was $R = -0.79$. When the cats under the age of 1-years were excluded from the analysis, as we know that younger cats are unusually active, the Pearson's correlation coefficient $R = -0.62$, described a strong negative correlation, see fig. 5.3.

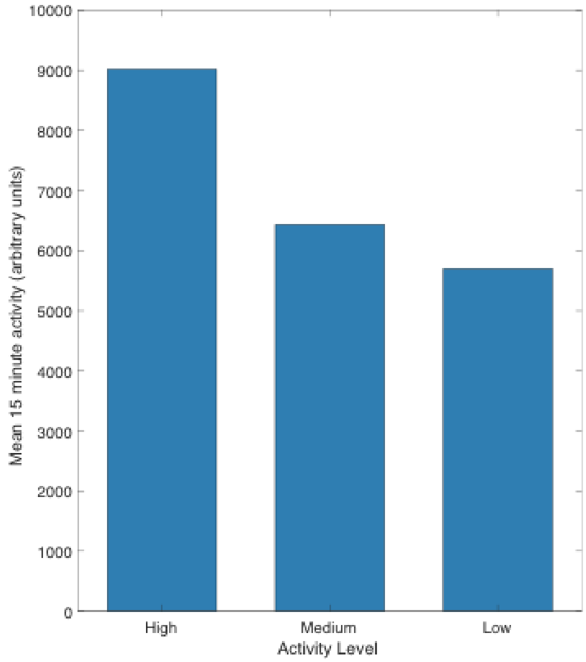


Figure 5.2: A comparison of the mean activity levels for the cats in categories of activity level as described by the technicians

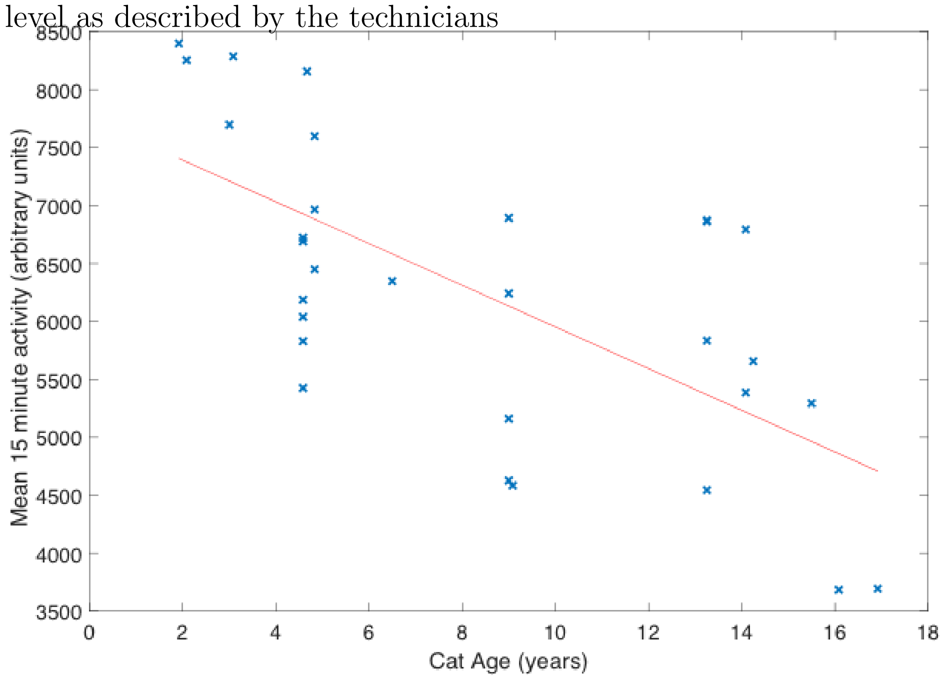


Figure 5.3: Mean activity taken over 15-minute periods plotted against the age of all cats older than 1 year. The red line marks a first-degree polynomial fit to the data and describes the negative trend.

5.3.1.3 Activity correlation with Body Condition Score

Due to the low sample size for the underweight and obese groups it is not possible to perform a statistical evaluation of the difference between these groups with confidence. Preliminary analysis of the difference between the overweight and ideal weight cats using an independent samples t-test was not significantly different at the 0.05 level (see fig. 5.4).

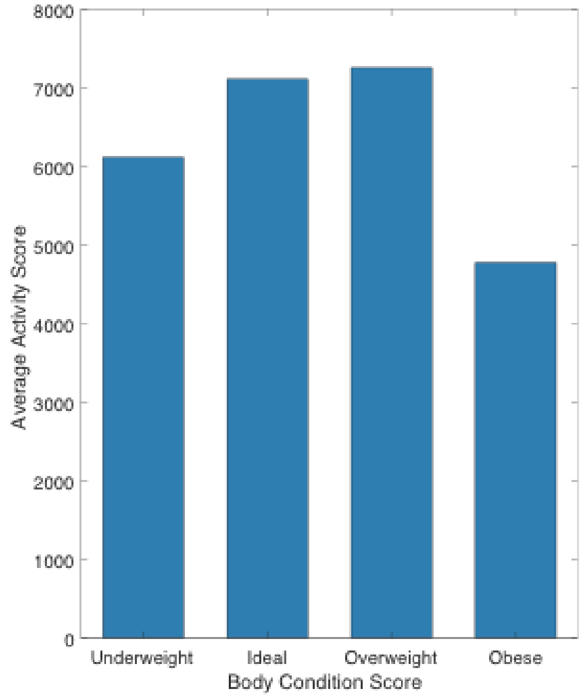


Figure 5.4: A comparison of the average activity level for cats in different body condition score categories.

5.3.1.4 Kitten activity

An independent samples t-test was performed to determine the difference in activity levels between kittens (cats less than 6 months old) and healthy adults. The difference was found to be statistically significant ($p < 0.001$) (see fig. 5.5).

5.3.1.5 Activity profiles

Activity profiles were generated for three cats from different subgroups of the study animals. Activity data were aggregated into 15-minute timeframes and plotted against time for a single 24-hour period.

The three profiles show distinctly different activity patterns in the cats. fig. 5.6a) shows the profile for a healthy adult male, fig. 5.6b) shows the profile for a geriatric female with a history of heart conditions and fig. 5.6c) shows the profile for a healthy female kitten. Whilst all three cats show clear peaks in activity, the geriatric female (fig. 5.6b) shows much fewer peaks and reaching a lower threshold, indicative of

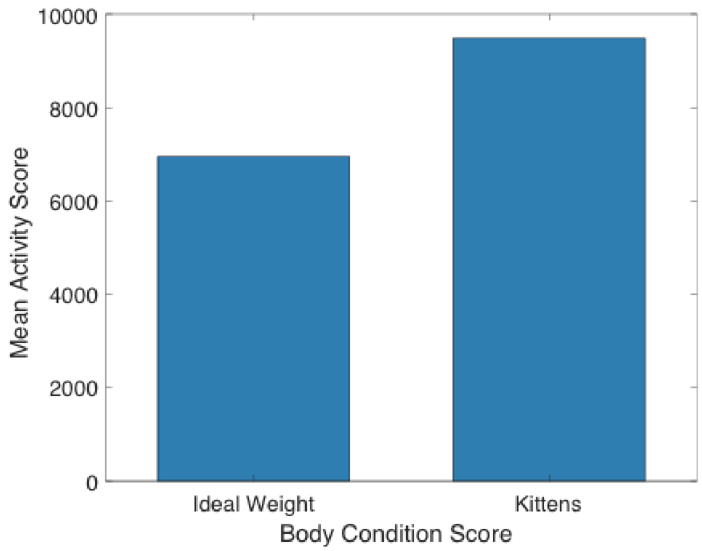


Figure 5.5: A comparison of the activity level of adult cats with an ideal body condition score and cats six months old and younger.

generally reduced activity throughout the day and less intense bouts of activity during active times when compared to the healthy male. The kitten (fig. 5.6c) shows sustained and intense peaks of activity throughout the day, with only short periods of inactivity during the late afternoon and very early in the morning.

Visualisations of median activity values for each 15-minute timeframe in a 24-hour period were calculated for all cats across all days (see fig. 5.7) and subsequently smoothed to remove noise using a Savitzky-Golay filter (see 5.8).

This visualisation shows a clear cycle present in the activity level of the cats. We can see that there are several large peaks in the daily pattern, followed by rapid declines with a secondary, lesser, peak two to three hours later, suggesting a period of high activity followed by a period of more sedate behaviour. Secondly, we can see that these peaks occur approximately every five hours starting at 1am and continuing through until approximately 9pm.

5.3.2 Study 2: Behaviour classification

F_1 scores were calculated for each class individually by first calculating the precision and recall values based on the output of the classifier predictions, see Table 5.3. The F_1 scores show strong classification performance for the movement-based behaviours, and weaker performance for the stationary behaviours. The mean F_1 score for the classifier across all classes, excluding the unspecified class is 0.729.

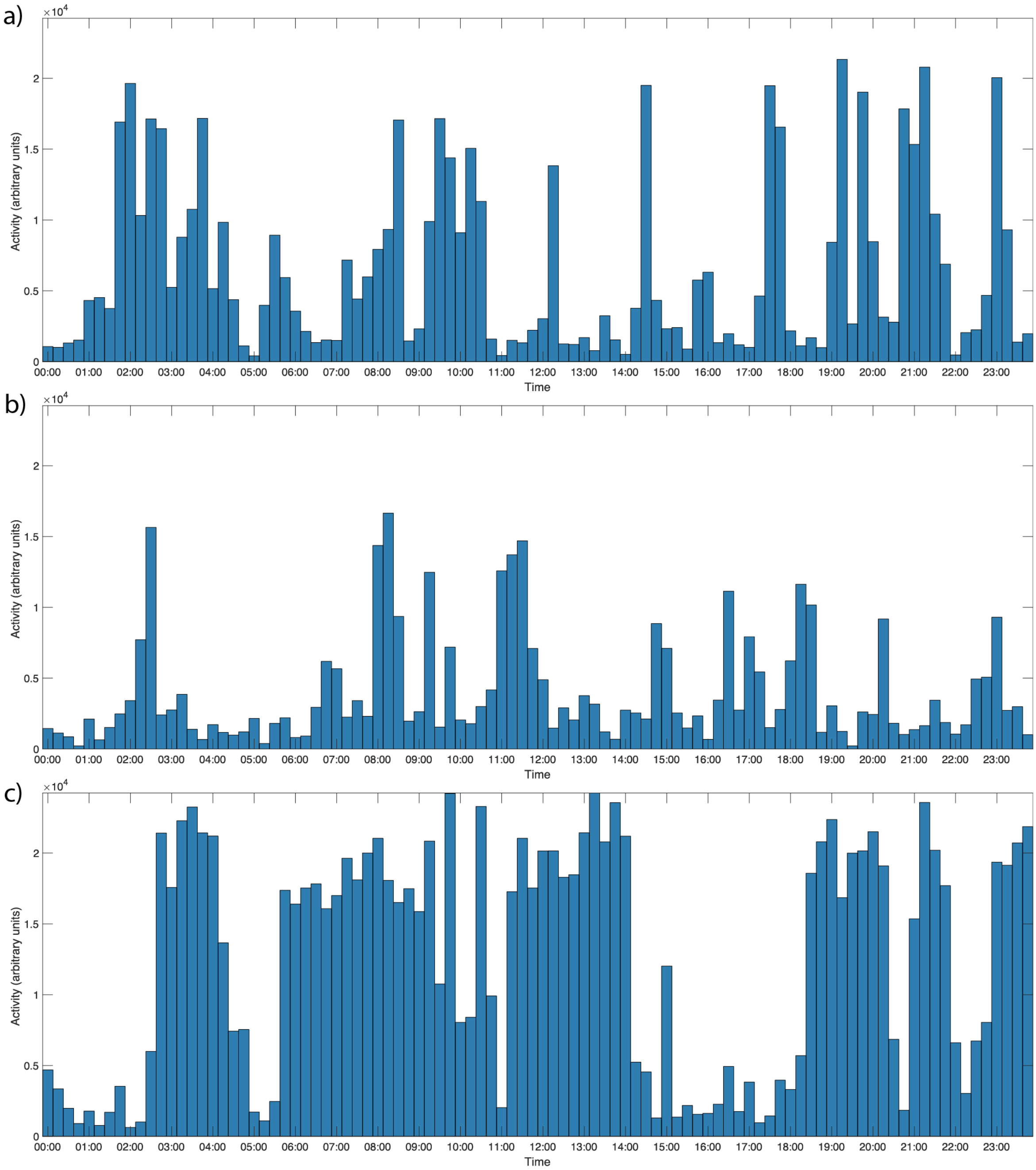


Figure 5.6: Activity profiles for three cats of differing ages in a single day. Activity was calculated from accelerometer data recorded by a collar mounted sensor. The three profiles show very different lifestyles associated with the age and health status of the cat. a) 6.5-year-old healthy male b) 16-year-old female with history of heart conditions, c) 6-month-old healthy female.

In order to visualise the performance of the classifier a confusion matrix has been produced, see fig. 5.9. The confusion matrix highlights areas in which the classifier performs less well. It is notable that there is substantial confusion between the eliminating behaviours (defecating and urinating) and the postures (sitting and lying). Interestingly sitting and standing have high recall values and low precision values,

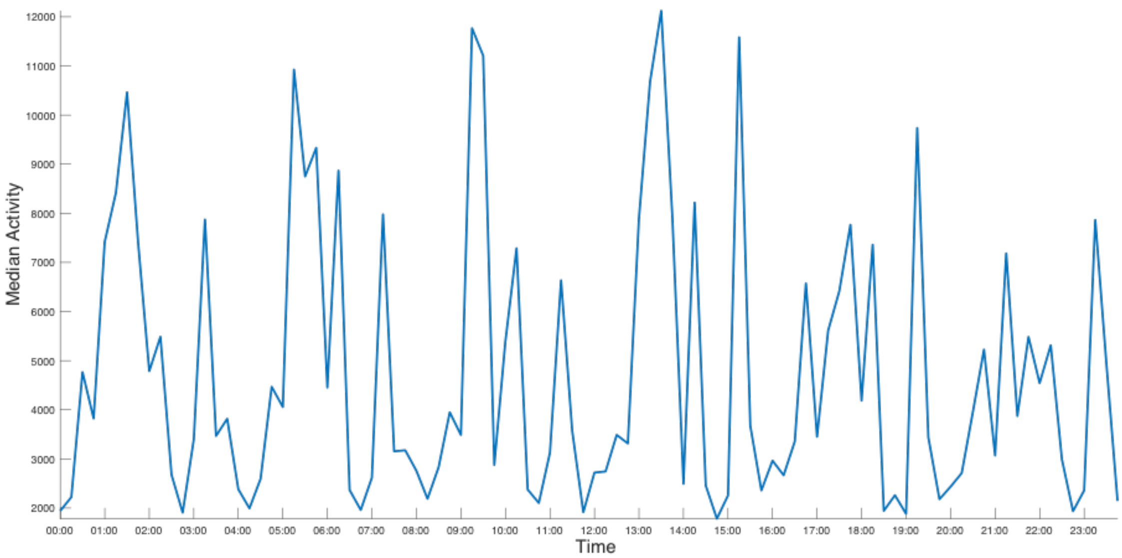


Figure 5.7: Median activity across all cats for each 15-minute window in each day.

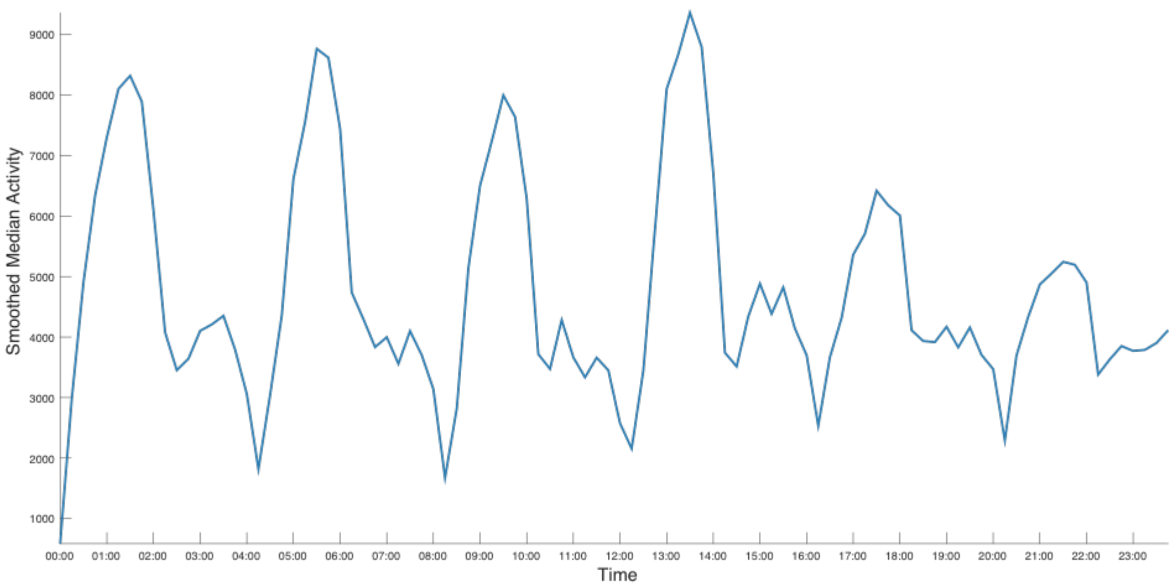


Figure 5.8: Median activity across all cats for each 15-minute window in each day, smoothed with a Savitzky-Golay filter to reveal underlying trends.

Table 5.3: Classifier performance metrics for the behaviours under analysis.

Class	Precision	Recall	F_1
Unspecified	0.267	0.181	0.216
Running	0.790	0.881	0.833
Walking	0.888	0.878	0.883
Jumping Up	0.854	0.757	0.802
Jumping Down	0.819	0.775	0.797
Scratching	0.844	0.769	0.805
Sitting	0.548	0.859	0.669
Lying	0.491	0.841	0.62
Eating	0.870	0.600	0.711
Drinking	0.670	0.700	0.685
Defecating	0.737	0.577	0.647
Urinating	0.728	0.467	0.569

suggesting that the low F_1 score for those behaviours is driven by the misclassification of the excreting behaviours.

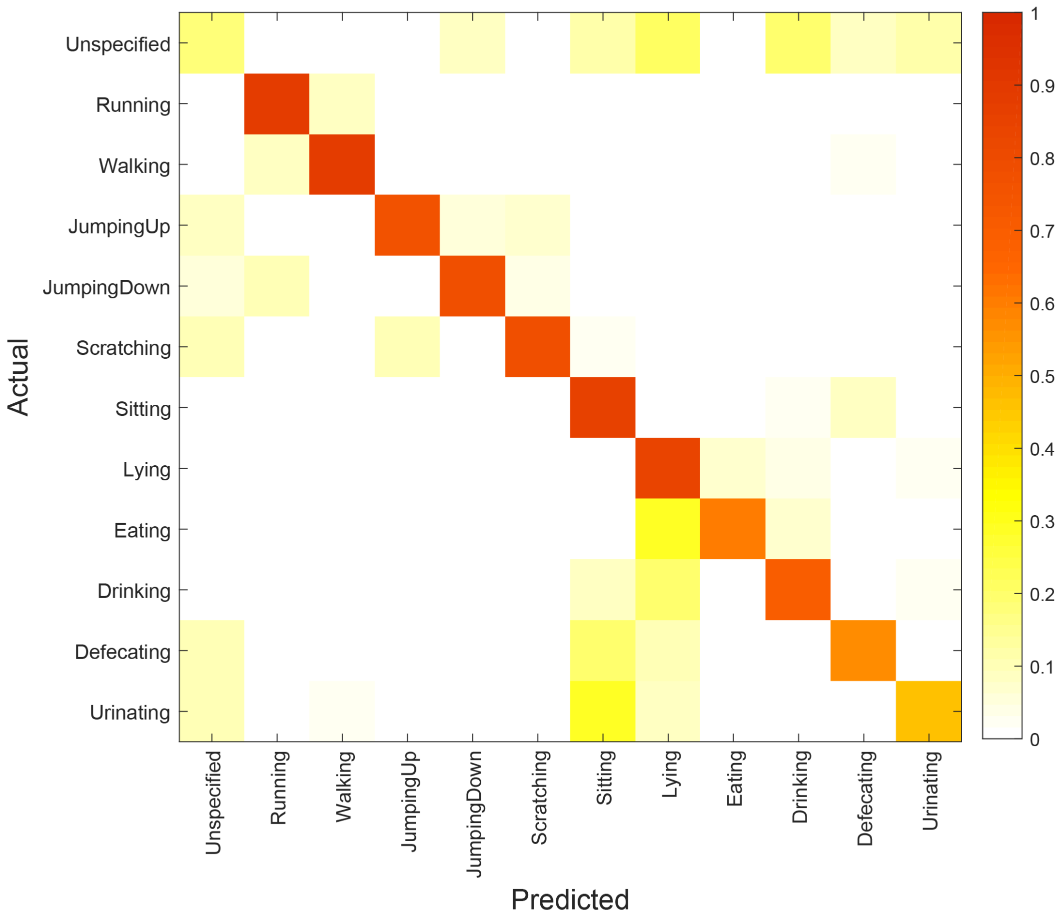


Figure 5.9: Confusion matrix outlining the results of predicting feline behaviour using a Support Vector Machine (SVM) classifier. The colour represents the proportions of the class that were predicted in relation to the number of instances of that class, a darker red is considered a better result. A strong dark series of cells diagonally across the matrix reflects accurate classification, as these cells show the proportion of correctly predicted windows. Darker cells outside of the main diagonal describe a higher proportion of windows misclassified.

5.4 Discussion

In this chapter we set out to develop a set of tools to facilitate the monitoring of feline behaviour using collar worn accelerometers. This was performed using data from two distinctly different scenarios. The first, in which colony-based, group housed cats were monitored, provides us insight into the general activity of cats, in a relatively constrained setting with fixed routines. The development of a behaviour recognition algorithm required annotated video data to provide ground truth for classifier training. Due to the length of this study a second data collection period, in which subjects were filmed whilst wearing sensors, was required as annotation of the first dataset would have been impractical.

The first study presented us with an opportunity to validate the performance of the sensors' measurements when compared to the subjective assessments of individuals familiar with the cats. We found that the technicians providing the assessment were confident in dividing the cats into three general activity levels: high, medium and low. When considering the aggregated activity of the cats we found that those cats described as being highly active by the technicians produced significantly higher sensor values than those described as medium, or low activity cats. Whilst the cats described as medium active were not significantly more active than the low activity cats according to the sensor ($p = 0.147$), the average sensor measured activity level was higher. This generally supports our assertion that the sensor is a reasonable proxy for the subjective assessment of the colony technicians.

We have shown that the age of a cat has a significant influence on the amount of activity it exhibits, especially when considering those cats under the age of six months. When performing a correlation analysis on the mean 15-minute activity of all cats we found a very strong negative correlation with age ($R = -0.79$). When excluding the kittens from this analysis the strength of the correlation was reduced ($R = -0.62$) whilst remaining strong. This makes intuitive sense, as we would expect that as a cat ages its demeanour becomes calmer continuing to do so into geriatric ages when the medical health of the cat begins to impose its own demands for inactivity [23]. This analysis demonstrates that the accelerometry is an appropriate tool for measuring the activity levels of cats, and could be used to identify cats that are behaving in an abnormal manner, signalling that a health assessment could be in order.

Body Condition Score (BCS) describes the physique of the cat, independently of gender or age, and as such has the potential to provide insight into the physical fitness of the cat [84]. Currently there are concerns regarding obesity in cats [12], by investigating how activity levels relate to the weight of the cat, insight into how to encourage healthier cats may be obtained. We performed an analysis of the impact of BCS on the activity level of the cat, as measured by the sensor. Whilst we did not have sufficient statistical power to confidently assess the significance of the difference between the groups, a visual assessment of the data would suggest cats "obese" cats are substantially less active than others. We would similarly expect to see that "overweight" cats would be less active than those of an ideal weight, however this is not borne out in the data and raises an interesting question regarding the mechanism used to measure activity. As the sensors measure the acceleration of the cat it is possible that bulkier cats produce more acceleration, driven by their larger size and increased amount of body fat. This could cause a larger amount of movement in the sensor than that produced by leaner cats, and consequently register higher activity values.

As described in the methods section, we excluded kittens (cats under the age of six months) from this analysis as BCS is not considered to be a useful measurement whilst the cats body is still developing. An analysis of the difference in activity levels between kittens and healthy adults, however, demonstrated the drastically different activity levels of the groups. This is reflected in the subjective assessments of the technicians, in addition to anecdotal evidence of the energetic nature of the kittens.

A commonly noted topic of interest raised through discussions with the technicians focused on the activity of the cats whilst the technicians were not in attendance, particularly through the night when the colony was minimally staffed. To this end we produced visualisations of the cats' activity on a day to day basis, see 5.6. The cyclic nature of these profiles, showing peaks in activity followed by resting periods led us to examine how the cycles might be carried across when considering the dataset as a whole. To visualise this, activity data was aggregated for all cats and median values were taken for each 15-minute period in a specific day. The resulting chart (fig. 5.7 showed a noisy, but clear cycle of activity in which a large peak in activity was followed by a dip, then a smaller peak, and then another dip, repeating approximately every 5 hours. Through discussion with the technicians and other staff we aimed to establish a reason for this underlying pattern. Feeding times were considered a possibility, whilst the cats were fed ad-libitum, the offer of fresh food can influence the rhythm of feeding behaviour [138]. Interactions with colony staff were considered not to be a factor as the pattern continues even when the colony was lightly staffed. Cats are known to have an activity rhythm that revolves around the day night cycle [23], however study of this rhythm at the resolution presented in this work has not been conducted.

The second study in this chapter aimed to produce a classification pipeline for the analysis of specific behaviours exhibited by cats in a domestic setting. Towards this, a dataset containing 31 cats exhibiting everyday behaviours whilst wearing collar-mounted sensors was collected. We found that the classifier performed well for behaviours that involved movement of the cat and less well for those behaviours in which the cat was stationary. The mean F_1 score for the classifier was 0.729 indicating a relatively good performance, although there is clearly room for improvement. The postures (sitting and lying) have relatively low F_1 scores (0.669 and 0.62 respectively), despite high recall values (0.859 and 0.841 respectively). This is driven by the classifier incorrectly predicting other behaviours as being one of the postures, specifically the eliminating behaviours (defecating and urinating). This confusion arises from the static nature of the cat during elimination. As the cat is in a stationary posture, there is little to differentiate it from instances when the cat is in a similar posture, but not performing elimination behaviour.

The algorithm performed less well when classifying behaviours in which the subject was stationary whilst exhibiting the behaviour, namely feeding, drinking, defecating, and urinating, and performed best when the subject was moving in consistent fashion (walking, running, jumping) or where the subject's posture was descriptive of the behaviour (sitting, lying, eating). Defecation and urination provided the highest degree of confusion for the classifier, a case that has been found to be true in similar studies on dogs [82]. The classifier developed for dogs by Ladha et al. [82] performed well for eating and drinking, however the physiology of the dog requires a grosser movement of the head in the exhibition of these behaviours which may have been easier to distinguish through accelerometry. Watanabe et al. also found success in their attempts to classify eating and drinking [149], however the approach taken in their work used posture classification to determine the behaviour and did not attempt to classify postures such as lying or sitting. Consequently, there was no possibility for misclassification between said postures and the eating/drinking behaviour, which is the primary source for confusion in our implementation. A potential approach to improving classification accuracies for these behaviours would be to consider that those behaviours for which the classifier underperformed are generally performed in a specific location. Eating and drinking occur at the food and water bowl respectively. Urination and defecation, at least for indoor cats, take place in the litter tray. The addition of location sensors, such as Bluetooth beacons, infrared sensors, or augmentations to food bowls/litter trays, would provide an additional layer of data which would be able to supplement the classification algorithm improve classification accuracy. There would still be cases in which this would not be sufficient, excretion outside of the litter tray for example, and further work should be conducted towards improving the classification based on accelerometry for these cases.

Assuming improvements to the classification accuracies are attainable, an analysis of the dataset produced for the activity study has the potential to provide further insight into the behaviour of colony cats, allowing for associations between behaviour and the health and age status of the cats to be established.

As described in the introduction to this chapter, the application of behaviour classification algorithms has been demonstrated for dogs, in the context of monitoring for health and wellbeing, as well as for satisfying curious owners. Further challenges are presented when considering this approach in the context of feline health assessment, particularly the increased agility of cats, reluctance to display sickness behaviours and the differing behavioural profiles associated with indoor and outdoor cats. The behaviour recognition algorithm's already strong performance in classification of locomotory behaviours (walking, running, jumping) would suggest suitability for assessment of degenerative joint diseases such as osteoarthritis and other movement impairing conditions. This kind of assessment could provide veterinarians with a further layer of information with which to diagnose and monitor the health of their patients, allowing for a continuous update on the mobility of the cat, and its reaction to various treatment methods.

5.5 Conclusion

In this paper we have demonstrated the ability of accelerometer data to characterise the behaviour of cats in a group housed setting. We have outlined the changes in activity that take place as cats age as well as providing insight into the daily rhythms of cat behaviour. We have developed a system for the classification of feline behaviours deemed to be of interest to owners and veterinarians and highlighted the challenges faced by such a system, as well as opportunities to exploit the technology to improve the welfare of the cats. The desire for descriptive data relating to the lives of our pets is burgeoning in increasingly data-centric world. Whether this desire is motivated by health and wellbeing reasons, or by curiosity about the lives of our pets, we have shown that an automated system presents an opportunity to capture this information in a detailed and reliable manner, allowing for interpretation by veterinarians and the potential to foster a deeper understanding of our pets.

6
General Discussion

Contents

6.1 Introduction

In the Introduction we highlighted the benefits of automated behaviour monitoring and assessment in livestock and companion animals. We discussed the improvements to productivity, welfare, and understanding of animal needs that such automation can facilitate. We have developed methods for automatically monitoring animal behaviour in three distinct species of domestic animals, and in four different contexts. As such it would be intuitive to think that we would require three entirely different approaches, and whilst this is true to a degree, we have shown that there is a common pipeline for algorithm development that is applicable across all of these scenarios. By focussing our approach to behaviour monitoring on the use of accelerometry (with the addition of gryoscopy in the case of Chapter 2), we have been able to develop and refine this pipeline towards reliably recording, and automatically classifying the exhibition of a diverse range of behaviours. Dependant on the application at hand, the classification algorithm was tailored to suit the behaviours under assessment. In each case-study, preliminary work was conducted to ensure that the movements classified would provide us with valuable information. The use of domain specific knowledge to identify behavioural patterns and activities was crucial to this process. In Chapter 2 we developed techniques for observing the movements of horses under the command of human riders, allowing us to automatically distinguish between the gait in which the horse was being ridden, and whether the horse was turning or travelling in a straight line. In Chapter 3 we examined the lying behaviour of farrowing sows, automatically classifying standing, sitting and lying postures, and expanded this in Chapter 4 by including behaviours that the sows were able to exhibit due to the removal of the constraints imposed on their movements by the farrowing crates. Finally, in Chapter 5 we developed algorithms to characterise the activity level of cats, and to automatically classify a range of behaviours deemed to be of interest to owners and veterinarians. Each of these scenarios involved the analysis of time-series data using signal processing and machine learning approaches. Through this research we have encountered challenges associated with the use of wearable sensors, but we have also been able to pinpoint key benefits.

6.2 Benefits of Automated Monitoring

Given that we have demonstrated, through validation and evaluation, that these approaches to activity classification are applicable across species and provide good levels of accuracy, it is necessary to draw comparisons between automatic and traditional methods of observing animal behaviour. All of the monitoring performed through the use of acceleration data could have been replicated by a dedicated human observer, however there are key considerations to be made when drawing this comparison. Perhaps the most important is the reduction of human interaction the approaches affords us. Through the work presented in chapters 3, 4, and 5 we have demonstrated the benefits of sensor-based approaches to monitoring animal behaviour in long-term studies. In Chapters 3 and 4 the sows were fitted with sensors which remained in place, and recording, for a period of nearly a week. In Chapter 5, the first study involved recording the activity of the cats for 2 weeks. Long-term observation of behaviour would not be possible without a team of observers, and even given that, continuous recording would be expensive, tedious and prone to inaccuracies. The sensor-based approaches outlined in this thesis each required an initial investment of time and planning in order to set the sensors recording and secure them to the subjects, in addition to the setup of recording hardware to allow for manual annotation of the video. Manual interaction with the sensors was minimal, and observation could have proceeded indefinitely if not for constraints of battery power and the need to manually extract the data from the sensors. As power usage and battery size reduce, we should expect to achieve longer studies, and as wireless technology allows us to remotely access the data recorded, manual intervention will be reduced significantly.

Another key benefit of sensor-based approaches to monitoring animal behaviour is the ability to record the movement of an animal at a very fine resolution. High sampling rates allow us to record changes in behaviour that last a fraction of a second, something that human observers would require slow motion video footage to be able to do. In Chapter 2 we were able to measure the timings of the footfalls of the horse within a few hundredths of a second. In Chapters 3 and 4 we were able to record the duration of postural transitions with high precision and use this as a factor to describe the control exhibited by the sow. The high resolution of the sensors (between 30 and 100Hz) also allowed us to examine specific instances of behaviour exhibition in fine detail. This allowed us to develop techniques for movement characteristic assessment that would not have been possible through manual observations without the use of sophisticated video capture devices.

6.3 Challenges Associated with Wearable Sensors

Despite the benefits detailed above, there are equally some drawbacks to the use of wearable sensor-based approaches to behaviour monitoring. Wearable sensors, by definition, require that the sensor be physically attached to the subject, presenting several challenges. In Chapter 2 we chose to secure the sensors to the legs of the horse, providing us with detailed data regarding the gait of the horse. This approach would have certainly been improved had we been able to also secure sensors to the saddle and noseband of the horse, however we found that riders were reluctant to attach a sensor to the expensive saddles, and the horse would not tolerate a sensor in its eye-line.

In Chapters 3 and 4 securing the sensors to the sow presented a significant challenge. Other researchers approaching this challenge had opted for collar [31] , or ear tag mounted sensors [111], however there are trade-offs associated with both approaches, namely noise in the signal from ear movements, and interference and rotation of the collar. Several different approaches were suggested before we concluded that, were we to record data from the hind end of the sow, strong adhesive and multiple layers of tape would be required. Even given this degree of reinforcement we still suffered from sensor removal due to the inhospitable environments in the farrowing crates and pens. The study in Chapter 5 required that we secure sensors to the collars of cats. Whilst we found that, following a period of habituation, the cats grew to tolerate the sensors; we were required to develop a technique to specifically identify the signal caused by cats scratching at the collar in an effort to remove it. The imposition of a sensor onto the body of an animal must be carefully considered before selecting wearable sensors as a suitable approach for monitoring behaviour. In each of the case-studies presented in this thesis considerable research was performed to ensure that the sensors would be as unobtrusive as possible to the animal, see Section 1.3.

Whilst not a limiting factor in the studies described in this thesis, the battery life of wearable sensors is a key consideration for any longer-term deployment. The accelerometers used in this study were capable of recording data at 100Hz for approximately 1-2 weeks. Reducing the sampling rate to 30Hz extends this to more than a month. As we have discussed in previous chapters, determining the optimal sampling rate is vital to ensuring that the full detail of the behaviour is captured. It is often acceptable to sample at a higher rate initially, and retrospectively down-sample the data to improve computational resource use. This approach is only suitable when the study is known to be of short enough duration to ensure the battery will last. Another option for preserving battery life would be to opportunistically turn the sensors on and off under certain conditions. For example, when a sensor attached to an animal has produced no signal for a predetermined period of time, it could be assumed that the animal is sleeping. Turning off the sensor in this situation, and turning it on again when the animal moves, would allow for long periods of time in which nothing of note is occurring to be removed from consideration, and increase battery duration.

6.4 Positioning and Use of Multiple Sensors

The capacity of accelerometers to accurately measure the movements of the subject has allowed us round-the-clock monitoring of movement, and as a result interpretation of behaviour. The choices surrounding the practical use of the sensors were the subject of substantial thought and research. There are several considerations to be made when using accelerometers, however we identified two as particularly influential: i) the positioning of the sensor on the subject, and ii) the number of sensors used. In each chapter in this thesis we have taken different approaches to these points. In Chapter 2 we used four sensors, one on each leg of the horse. It was felt, given that we were aiming to assess the gait of the horse, that this would provide the most relevant data. By securing the sensors to the legs of the horses we were able to determine the footfalls of each leg, allowing for assessment of regularity, rhythm and duty factor. Gait recognition on the other hand was performed only using data from a single sensor on one of the horse's front legs. The balance here is to capture as much relevant data as possible whilst ensuring minimal imposition on the horse's behaviour. As described briefly above, more sensors were considered for use in this study, however were ruled out due to the impact on the horse and rider. A key contribution of Chapter 4 was a comparison of the sensor positioning and numbers when applied to lying behaviour assessment. We found that the addition of a second sensor dramatically increased the performance of the classification algorithm. Similarly, the position of the sensor on the animal had a substantial impact, with the hind sensor performing better than the front sensor when both were used independently. This allows us to infer the importance of sensor positioning in this context and speaks more broadly of the relevance of this as a consideration when designing future studies. Finally, in Chapter 5, we used a single sensor mounted to the collar of the subject. This was deemed to be the only possible location for the sensor without requiring a much more invasive attachment.

As mentioned, there is a balance to be drawn when considering the practicality and feasibility of multiple sensors. Each additional sensor increases the amount of data collected and can improve the conclusions that we are able to draw as a result. This comes at the cost of increased computational requirements, imposition to the subject and exposure to hardware failure. The need for large amounts of computational power have been evident throughout this project. With data sets totalling hundreds of gigabytes, training times for the classifiers often extended over many days. Hardware failure is a problem that was faced in all of the studies. In some cases, we were able to account for this by building in redundancy, as in Chapter 2 in which each sensor had a backup, however in some instances this lead to a loss of data, as in Chapter 5. As hardware standards improve, it is reasonable to assume that this will become less of a problem. However, even with perfect hardware, when delicate electronics are introduced into such inhospitable environments as a farrowing pen, failure is inevitable.

6.5 Movement Characteristic Assessment

The first computational task in each of the studies described herein has been the development of an activity recognition algorithm, allowing us to identify the specific activity that is being exhibited by the subject. Whilst this provides us with an approach to monitor the behaviour of the animal, it is possible to learn more about the animals by not only monitoring the behaviours they exhibit, but also considering the characteristics of each specific exhibition. As described above, the high sample rate afforded to us by the sensors allowed for the examination of instances of behaviour exhibition in minute detail, providing a level of information that is not easily accessible through manual observation. By extracting salient characteristics of a behaviour from the time-series data, we are able to compare two events of the same kind and draw inferences about the animals under observation. This has been described in Chapter 2, in which we used movement characteristic assessment as a proxy for determining the skill with which a rider's commands were executed by the horse. Through consultation with experts we were able to target aspects of the data that best represented a range of criteria upon which horses are traditionally judged. Whilst this is a task that is traditionally performed manually (by judges and coaches) the sensors allowed us to remove the dependence on manual observation and provided objective feedback on the performance of the horse and rider. In Chapters 3 and 4, by defining a set of characteristics that were determined to be reflective of a controlled posture transition, we were again able to find aspects of the accelerometry data to describe this. Some of these were intuitive, such as the duration of a lying event, however others were less so. The "jerk" of a transition, for example, describes the first derivative of acceleration, we determined this to relate to the smoothness of the transition, and subsequently the control exhibited by the sow.

The process of subjectively identifying movement characteristics that are suitable for emulation in the accelerometer data relies upon access to an expert. The provision of domain specific knowledge is an essential component of each of the studies included in this thesis. This however comes with drawbacks. Reliance upon subjective assessments depend upon the skill and experience of the initial assessor. This problem was evident during the assessment phase of Chapter 2. Three judges were recruited to provide assessments of the horses' performances, against which we hoped to compare our metrics. By performing an inter-rater reliability assessment on the judges' assessments however we found that the judges not only did not agree with our assessments, but also did not agree with each other. This is by no means an indictment of the skill of the judges, but rather speaks to the complexity of this kind of problem. The problem of subjective assessments differing between assessors in well known. Research into lameness assessment in horses has found that scores differed, even between experts [72]. Whilst there are approaches to quantify the reliability of scorers [137], it is reasonable that we should be aiming to remove subjective assessment altogether. The research presented in this thesis presents a step towards this in some fields.

With the ongoing development of deep learning techniques there is a movement away from the necessity of domain specific knowledge. The accessibility of increasingly large datasets has meant that rather than applying domain specific knowledge to identify aspects of the data salient to the task, deep neural networks can be trained on raw data alone. Whilst this certainly has benefits, there is a balance to be drawn between results and efficiency, and intuitiveness. Where a deep learning algorithm may be able to determine the skill of a horse and rider pair, for example, the specific characteristics of a dressage performance that describe said skill would be particularly difficult to determine. Through the approaches developed in this thesis, that intuition remains intact. By explicitly defining the characteristics of the movement that describe the outcome we are aiming for, we are better able to report back to the interested party, informing future decisions regarding the handling of the animal.

6.6 Impact of Behavioural Complexity

This thesis aimed to describe the challenges faced when developing wearable sensing solutions towards monitoring and assessing animal behaviour. A key aspect of this has been to describe how these technical challenges change as the complexity of the scenario increases. As described previously we have aimed to describe a clear progression of complexity in terms of freedom of the animal to perform natural behaviours. In Chapter 2 we describe a scenario where the subject has little or no freedom to behave naturally. The riders were issued with a predetermined test describing the movements the horse was to complete. This allowed us to focus specifically on the classification of a very limited set of movements, namely the gaits and turns. Through this restriction we were able to produce a very specialised algorithm for delivering feedback to the rider, without the requirement for segmenting out extraneous behaviours.

In Chapter 3 we examined the behaviour of sows when confined to farrowing crates. The movement of the sows was considerably restricted, allowing them only the freedom to stand and lie. This afforded us the ability to target posture transitions with the aim of developing a framework for the assessment of lying behaviour. By performing a similar study on sows housed in free-farrowing environments (Chapter 4) we faced increased challenges in the form of a larger range of natural behaviour that the sows were able to exhibit. By its nature, this increased the complexity of the solution that was required to classify the activity of the sows, introducing more classes and more transition types.

In Chapter 5 we examine an essentially constraint-free environment. In the first study of the feline project we examined group housed cats whose behaviour was entirely at their own discretion, taking aside that they were housed indoors. The second study in Chapter 5 examines the challenges faced when developing an activity recognition algorithm for highly agile and fast-moving animals, as well as introducing compound behaviours such as play. This progression in complexity has allowed for a systematic approach to the development of algorithms for the assessment of animal behaviour.

Lessons learned in the initial work were taken into account in the study design and development stages of the later studies. In Chapter 2 we were not required to segment salient behaviours given that we would assess the full ride. In Chapter 3, by segmenting the data initially based upon a heuristic approach we were able to target specifically the transition periods between postures. In Chapter 4 we expand upon this and use this pre-segmented data as a platform to perform transition classification in addition to posture/activity recognition.

Whilst we see a clear progression of the complexity of the activity recognition task through the chapters, the same does not apply to the movement characteristic assessment, in which the complexity is dictated by the task. It would be easy to argue that the assessment performed in Chapter 2 is the most complex. The subjective nature of dressage assessment meant there were a large number of variables to consider when delivering feedback, and as a result we were required to simplify these in order to provide relevant information to the rider. It could be argued that our assessments suffered due to this, however it was felt that giving consistent objective feedback that was actionable by a rider was more important than strictly adhering to the subjective guidelines outlined by the judges. The assessments performed in Chapters 3 and 4 on the other hand, were more focussed on the kinematic properties of the movements exhibited and as a result had close relation to the acceleration data. This does not detract from the usefulness of these assessments however, rather it highlights the requirements for a tailored approach, and the ongoing need for domain specific expertise. The assessments performed in Chapter 5 were, possibly, the most simplistic, despite the highest degree of complexity in the activity recognition task. By analysing the activity level of the cats, full interpretation of behaviour is not leveraged from the data, however given the large number of behaviours under observation it was felt that at this point a more simplistic analysis could provide deeper insight into the patterns of behaviour exhibited by the population of cats.

6.7 Research Limitations

Developing tools that are applicable across multiple species and in various different contexts presents a range of complications that would have been more readily addressed if the scope had been narrower. There have been several limitations to the research and, given the benefit of hindsight, routes to address these should be considered.

A primary limitation of the work in this thesis revolves around sample size, particularly in Chapters 3 and 4. In both cases, sample size was limited due to the availability of sows due to farrow, time constraints and data loss due to hardware failure. Whilst for the purposes of demonstrating the applicability of the technology to the scenarios the sample size was sufficient, a larger sample size, particularly when comparing groups of sows, would have certainly improved the robustness of the results. We have made every effort to ensure that the evaluation and validation phases of each study took into account variations in individuals, however increasing the amount of data improve the accuracy of the classification algorithms when exposed to data from previously unseen individuals. This would require longer periods of data collection in which more sows could be assessed and closer collaboration with farm staff to preselect sows.

The study outlined in Chapter 2, explored the use of wearable sensing to provide feedback to amateur riders. We made the decision relatively early to focus on riders at the lowest level of dressage, as we believed they would be the ones who would have the clearest faults in their performance, as well as being those who would have least access to costly trainers and time to devote to practice. This has limited the variety of movements that were displayed during the tests. Whilst this simplified the problem to a degree, as we were left with a smaller set of activities to monitor, it imposes a limitation on the usefulness of the developed solution at any higher levels of performance. More experienced riders would almost certainly be aware that they were making the errors our algorithm was designed to highlight, and as a result would find little value in them. Addressing this would necessitate the development of activity recognition algorithms capable of recognising the full spectrum of dressage movements, and further consultation with judges to identify characteristics suitable for feedback.

The interdisciplinary nature of this research has pushed us to find a balance between addressing complex computational problems and developing novel animal science outcomes. We have attempted to find a middle ground and, as a result, certain areas of the research may be considered to lack the depth that one would expect if considering the problem from just one of these perspectives. Throughout this thesis we have employed the Support Vector Machine classification algorithm. This is a well understood algorithm that is known to perform well for a variety of applications. However, as mentioned previously, recently there has been movement in the Machine Learning community towards deep learning, which has shown exceptional performance in classification problems. Given more time and resources, an investigation into the use of deep learning, and specifically LSTM neural networks (Long Short-Term Memory) [134] for classifying data relating to animal behaviour could provide improvements to the results described herein. Conversely, in the feline studies we devoted more time to developing the classification algorithm, and as a result the outcomes regarding the interpretation of the cats' behaviour are less involved. This research is a very suitable target for extension, as the output of the classification algorithm could be used to gain insight into the health status of the cats at a level of detail that the activity monitoring alone was unable.

As described in the body of this thesis, a considerable amount of time and effort was dedicated to the annotation of video and sensor data in order to provide samples for training of supervised classification algorithms. A subset of classification algorithms that were not explored were unsupervised algorithms. The differences between supervised and unsupervised algorithms are apparent in the manner in which the classifier is trained. Where a supervised algorithm seeks to develop a model of the various classes under assessment through exposure to labelled examples, an unsupervised approach operates without having seen prior examples. Instead, unsupervised approaches look to find similarities between various groupings within the dataset, which can be retrospectively assigned to appropriate classes. These approaches can be particularly valueable for the analysis of very large datasets for which labelling is either impossible or impractical. Typically a subset of the data would be labelled to provide a platform for validation and to establish performance metrics. In the context of the work presented in this thesis, we deemed unsupervised approaches to be unnecessary as, despite the time commitments required, annotation was still feasible.

6.8 Applications Arising from This Work

The proof of concepts established herein present the foundation for further research and application development. This has already been acted upon in the case of the transition assessment algorithms developed in Chapter 3. A large-scale study involving more than 500 sows has been performed [100]. This study found associations between the kinematic characteristics of transitions and increased maternal crushing. In order to incorporate such traits into the breeding of livestock, we must demonstrate that such traits are heritable. In order to assess such heritabilities a very large number of animals and very clear pedigree information is required. Further utilisation of the other research described here may require more investment however.

Each of the scenarios presented herein have operated under a predefined protocol that involved a period of data collection and retrospective analysis and evaluation. In order for the tools we have developed to achieve widespread usability it is important to consider how we might apply them in a real time context. This is particularly relevant for the assessment of animal welfare, whether in the home or on the farm. The ability of wearable technology to assist in the identification of challenges to welfare, as in farrowing prediction for example, only becomes useful when it is able to provide notifications at the time of the event. There are several challenges to this however, first of which is the computational resources required to operate the algorithms. In several instances of the work in this thesis, steps were taken to ensure that processing would be kept as efficient as possible, however for a truly real time application this would have to be significantly improved. Equally, each of the tools developed would require integration with a networked system to allow for dissemination of the information generated. In the case of the work presented in Chapter 2, for example, it is easy to envision a mobile application which could deliver near real time feedback to a rider. This would require a team of developers with experience in mobile development, machine learning, and data architecture, but has the capacity to provide a solution for dressage training which is accessible to anyone with a horse and a smart phone.

Building upon the work presented in the feline study, the development of a real-time monitor for unwanted behaviour would allow for corrective measures to be taken. This could be achieved through integration of the algorithms with the output of other sensing modalities. It was suggested that pet owners would like to be aware of the amount of time their cat spent in out-of-bounds areas, such as furniture or on counter tops. In our work, we aimed to gain insight into this by recognising jumping up and down from surfaces, however a more involved approach using location sensors, such as Bluetooth beacons, could be used. Similarly, our approach struggled to accurately identify elimination behaviours. The use of sensors to determine when the cat was in the litter tray could allow us to refine this and build a larger data set with which to train the classifier by labelling elimination without the need for continuous observation. This could allow the recognition of these behaviours outside of the litter-tray and, in conjunction with a notification delivery service such as a smartphone, allow for owner intervention.

The techniques developed in this thesis have focussed on three species, however through this it has been demonstrated that these techniques are widely applicable given a degree of tailoring to the species in question. It is logical then to explore the possibility of exploiting this technology for the assessment of other species' behaviours. Significant work is already underway to commercialise the assessment of behaviour of cattle, dogs, horses and others using wearable technology. This is a rich field of research and investment into this, still relatively nascent, technology has the potential to leave a lasting impact on the way we understand the animals in our care.

6.9 Future research

During the course of this project, and based on the outcomes we have established, potential for further research has become apparent.

For the purposes of assessing the performance of dressage horses, we employed sensors secured to the legs of the horse. This allowed us detailed insight into the way the horses legs were moving, however we did not take into account the movements of the rider. The sensors we used were speciality devices, and as such relatively expensive. Given that the majority of riders are likely to carry smartphones with them, and the fact that modern phones have built in accelerometers, an investigation into performing dressage assessment on a smartphone could lead to a more accessible platform. We also limited ourselves to the assessment of dressage which is, in its nature, a restricted environment. Horses also engage in a range of other activities, such as eventing, show jumping, and hacking. Considering applications for wearable sensing to incorporate the assessment of these activities, has the potential to expand the use of these device and reach a wider audience in the horse-riding community.

The two porcine studies made strides to addressing the question of improving the welfare of sows and piglets in the periparturient period. A key next step would be to quantify the associations between maternal lying behaviour and welfare outcomes such as piglet mortality rates, nesting material requirements, and suitability for free-farrowing. The tools developed in this thesis would be suitable for use in establishing these, and given a large enough study, could have significant impact in both welfare and productivity of commercial pig units. Key , however, would be the consideration of how we objectively quantify welfare. Where it is possible to measure productivity in a numerical sense, welfare is less tangible. Research into establishing a metric for welfare that could be applied to pig units would allow us to quantify the impact our tools have, and lessen our dependence on subjective assessments.

Our investigations into cat behaviour raised some interesting questions. We found cyclic patterns in the cats' behaviour, and as we discussed these could be associated with feeding routines. The cats that took part in this study were, however, indoor housed cats. Considering how a cat's behaviour changes when given free access to outside areas could provide valuable insight into whether these patterns hold. Certainly, cats with access to the outside world would exhibit a range of behaviours in addition to those shown by house cats. Development of a classifier for all of these behaviours has the potential to allow us a much deeper understanding of the lives of our pets.

Bibliography

[1] Abowd, D., Dey, A. K., Orr, R. and Brotherton, J. [1998], 'Context-Awareness in Wearable and Ubiquitous Computing', *Virtual Reality* **3**(3), 200–211.

[2] Ahmadi, A., Rowlands, D. D. and James, D. A. [2006], Investigating the translational and rotational motion of the swing using accelerometers for athlete skill assessment, *in* 'Proceedings of IEEE Sensors', IEEE, pp. 980–983.

[3] Alcaidinho, J., Yoder, N., Mundell, P. and Valentin, G. [2014], 'Assessment of Working Dog Suitability from Quantimetric Data', *Georgia Institute of Technology* .

[4] American Horse Publications [n.d.], '2009-2010 AHP Equine Industry Survey Summary Statistics', online (02-2015): `https://www.americanhorsepubs.org/resources/AHP-Equine-Survey-Final.pdf`.

[5] Aparna, U., Juul Pedersen, L. and Jørgensen, E. [2014], 'Hidden phase-type Markov model for the prediction of onset of farrowing for loose-housed sows', *Computers and Electronics in Agriculture* **108**, 135–147.

[6] Atallah, L. and Yang, G.-Z. [2009], 'The use of pervasive sensing for behaviour profiling – a survey', *Pervasive and Mobile Computing* **5**(5), 447–464.

[7] Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R. and Havinga, P. [2010], Activity recognition using inertial sensing for healthcare, wellbeing and sports applica- tions: A survey, *in* '23rd Intl. Conf. on Architecture of Computing Systems (ARCS)', VDE, pp. 1–10.

[8] Axivity [2013], 'AX3: Triaxial Logging Accelerometer', Accessed: 26 Feburary 2018.
URL: *https://axivity.com/product/ax3*

[9] Axivity [2014], 'Axivity WAX9 - 9-Axis Bluetooth Streaming IMU', Accessed: 26 February 2018.
URL: *https://axivity.com/downloads/wax9*

[10] Bächlin, M., Förster, K. and Tröster, G. [2009], SwimMaster: a wearable assistant for swimmer, *in* 'Proceedings of the 11th international conference on Ubiquitous computing', ACM, pp. 215–224.

[11] Baxter, E. M., Jarvis, S., Sherwood, L., Farish, M., Roehe, R., Lawrence, A. B. and Edwards, S. A. [2011], 'Genetic and environmental effects on piglet survival and maternal behaviour of the farrowing sow', *Applied Animal Behaviour Science* **130**(1-2), 28–41.

[12] Beaver, B. [2007], The Welfare of Cats, *in* I. Rochlitz, ed., 'Anthrozoos', Vol. 20, Dordrecht: Springer, pp. 79–80.

[13] Benito, J., Hansen, B., Depuy, V., Davidson, G. S., Thomson, A., Simpson, W., Roe, S., Hardie, E. and Lascelles, B. D. [2013], 'Feline musculoskeletal pain index: Responsiveness and testing of criterion validity', *Journal of Veterinary Internal Medicine* **27**(3), 474–482.

[14] Blackshaw, J. K. and Hagelsø, A. M. [1990], 'Getting-up and lying-down behaviours of loose-housed sows and social contacts between sows and piglets during Day 1 and Day 8 after parturition', *Applied Animal Behaviour Science* **25**(1-2), 61–70.

[15] Blokhuis, H. J., Veissier, I., Miele, M. and Jones, B. [2010], 'The welfare quality® project and beyond: Safeguarding farm animal well-being', *Acta Agriculturae Scandinavica A: Animal Sciences* **60**(3), 129–140.

[16] Boers, M., Dijkmans, B. A., Breedveld, F. C., Camps, J. A., Chang, P. C., Brummelen, P. V., Pauwels, E. K. and Cats, A. [1990], 'Subclinical renal dysfunction in rheumatoid arthritis', *Arthritis & Rheumatism* **33**(1), 95–101.

[17] Brambell, F. [1965], 'Report of the Technical Committee to Enquire Into the Welfare of Animals Kept Under Intensive Livestock Husbandry Systems', Accessed: 26 Feburary 2018.
URL: *http://www.worldcat.org/title/report-of-the-technical-committee-to-enquire-into-the-welfare-of-animals-kept-under-intensive-livestock-husbandry-systems/oclc/475844139*

[18] British Horse Industry Confederation [2009], 'BHIC Briefing – Size and Scope of the Equine Sector', online (02-2015): www.bhic.co.uk/downloads/sizescope.pdf.

[19] Brugman, H., Russel, A. and Nijmegen, X. [2004], Annotating Multimedia/Multi-modal Resources with ELAN., *in* '4th International Conference on Language Resources and Language Evaluation, European Language Resources Association', pp. 2065–2068.

[20] Bulling, A., Blanke, U. and Schiele, B. [2014], 'A tutorial on human activity recognition using body-worn inertial sensors', *ACM Computing Surveys* **46**(3), 1–33.

[21] Cagnacci, F., Boitani, L., Powell, R. A. and Boyce, M. S. [2010], 'Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges', *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**(1550), 2157–2162.

[22] Cain, P. J., Guy, J. G., Seddon, Y., Baxter, E. M. and Edwards, S. E. [2013], 'Estimating the economic impact of the adoption of novel non-crate sow farrowing systems in the UK', *International Journal of Agricultural Management* **2**(2), 113–118.

[23] Casey, R. A., Bradshaw, J. W. S. and Brown, S. L. [2012], *The behaviour of the domestic cat*, number Book, Whole, 2nd edn, University of Bristol.

[24] Chapinal, N., de Passillé, A., Pastell, M., Hänninen, L., Munksgaard, L. and Rushen, J. [2011], 'Measurement of acceleration while walking as an automated method for gait assessment in dairy cattle', *Journal of Dairy Science* **94**(6), 2895–2901.

[25] Clayton, H. M. and Schamhardt, H. [2013], Measurement Techniques for Gait Analysis, *in* 'Equine Locomotion', London: Elsevier BV, pp. 31–60.

[26] Clayton, H. et al. [1990], Time-motion analysis in equestrian sports: the grand prix dressage test., *in* 'Proceedings of the Annual Convention of the American Association of Equine Practitioners', Vol. 35, pp. 367–373.

[27] Cleland, I., Kikhia, B., Nugent, C., Boytsov, A., Hallberg, J., Synnes, K., McClean, S. and Finlay, D. [2013], 'Optimal placement of accelerometers for the detection of everyday activities.', *Sensors* **13**(7), 9183–9200.

[28] Cochran, W. W. and Lord Jr., R. D. [1963], 'A Radio-Tracking System for Wild Animals', *The Journal of Wildlife Management* **27**(1), 9–24.

[29] Cornou, C. [2006], 'Automated oestrus detection methods in group housed sows: Review of the current methods and perspectives for development', *Livestock Science* **105**(1-3), 1–11.

[30] Cornou, C. and Kristensen, A. R. [2014], 'Monitoring individual activity before, during and after parturition using sensors for sows with and without straw amendment', *Livestock Science* **168**, 139–148.

[31] Cornou, C. and Lundbye-Christensen, S. [2008], 'Classifying sows' activity types from acceleration patterns. An application of the Multi-Process Kalman Filter', *Applied Animal Behaviour Science* **111**(3-4), 262–273.

[32] Cornou, C. and Lundbye-Christensen, S. [2010], 'Classification of sows' activity types from acceleration patterns using univariate and multivariate models', *Computers and Electronics in Agriculture* **72**(2), 53–60.

[33] Cornou, C. and Lundbye-Christensen, S. [2012], 'Modeling of sows diurnal activity pattern and detection of parturition using acceleration measurements', *Computers and Electronics in Agriculture* **80**, 97–104.

[34] Cornou, C., Lundbye-Christensen, S. and Kristensen, A. R. [2011], 'Modelling and monitoring sows' activity types in farrowing house using acceleration data', *Computers and Electronics in Agriculture* **76**(2), 316–324.

[35] Cronin, G. M. and Smith, J. A. [1992], 'Effects of accommodation type and straw bedding around parturition and during lactation on the behaviour of primiparous sows and survival and growth of piglets to weaning', *Applied Animal Behaviour Science* **33**(2-3), 191–208.

[36] Damm, B. I., Forkman, B. and Pedersen, L. J. [2005], 'Lying down and rolling behaviour in sows in relation to piglet crushing', *Applied Animal Behaviour Science* **90**(1), 3–20.

[37] Damm, B. I., Vestergaard, K. S., Schrøder-Petersen, D. L. and Ladewig, J. [2000], 'The effects of branches on prepartum nest building in gilts with access to straw', *Applied Animal Behaviour Science* **69**(2), 113–124.

[38] Davies, E. R. [2009], 'The application of machine vision to food and agriculture: a review', *The Imaging Science Journal* **57**(4), 197–217.

[39] Dawkins, M. S. [2004], 'Using animal behaviour to assess animal welfare', *Animal Welfare* **13**(S1), 3–7.

[40] Dayhoff, M. O. and Schwartz, R. M. [1978], A model of evolutionary change in proteins, *in* 'In Atlas of protein sequence and structure', Citeseer.

[41] Edwards, J. C. [1988], 'The effects of Trichinella spiralis infection on social interactions in mixed groups of infected and uninfected male mice', *Animal Behaviour* **36**(2), 529–540.

[42] Edwards, S. A. [2006], 'Tail biting in pigs: Understanding the intractable problem', *Veterinary Journal* **171**(2), 198–199.

[43] Edwards, S. A., Brett, M., Ison, S., Jack, M., Seddon, Y. M. and Baxter, E. M. [2012], Design principles and practical evaluation of the PgSAFE free farrowing pen, *in* '4$^{\text{th}}$ European Symposium on Porcine Health Management', p. 113.

[44] Escalante, H. J., Rodriguez, S. V., Cordero, J., Kristensen, A. R. and Cornou, C. [2013], 'Sow-activity classification from acceleration patterns: A machine learning approach', *Computers and Electronics in Agriculture* **93**, 17–26.

[45] Falvey, J. L. [1987], *An introduction to working animals*, Australia: MPW Australia.

[46] Fédération Equestre Internationale [2007], *Dressage Handbook Guidelines for Judging*, Fédération Equestre Internationale, Lausanne, Switzerland.

[47] Fraser, D. [1993], 'Assessing Animal Well-Being: Common Sense, Uncommon Science', *Food animal well-being* pp. 37–54.

[48] Gäde, S., Bennewitz, J., Kirchner, K., Looft, H., Knap, P. W., Thaller, G. and Kalm, E. [2008], 'Genetic parameters for maternal behaviour traits in sows', *Livestock Science* **114**(1), 31–41.

[49] Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M. and Toulmin, C. [2010], 'Food security: The challenge of feeding 9 billion people', *Science* **327**(5967), 812–818.

[50] González, L., Tolkamp, B., Coffey, M., Ferret, A. and Kyriazakis, I. [2008], 'Changes in Feeding Behavior as Possible Indicators for the Automatic Monitoring of Health Disorders in Dairy Cows', *Journal of Dairy Science* **91**(3), 1017–1028.

[51] Grandinson, K., Lund, M. S., Rydhmer, L. and Strandberg, E. [2002], 'Genetic parameters for the piglet mortality traits crushing, stillbirth and total mortality, and their relation to birth weight', *Acta Agriculturae Scandinavica - Animal Science* **52**(4), 167–173.

[52] Grandinson, K., Rydhmer, L., Strandberg, E. and Thodberg, K. [2003], 'Genetic analysis of on-farm tests of maternal behaviour in sows', *Livestock Production Science* **83**(2-3), 141–151.

[53] Grégoire, J., Bergeron, R., D'Allaire, S., Meunier-Salaün, M. C. and Devillers, N. [2013], 'Assessment of lameness in sows using gait, footprints, postural behaviour and foot lesion analysis', *Animal* **7**(7), 1163–1173.

[54] Gruen, M. E., Griffith, E. H., Thomson, A. E., Simpson, W. and Lascelles, B. D. X. [2015], 'Criterion validation testing of clinical metrology instruments for measuring degenerative joint disease associated mobility impairment in cats', *PLoS ONE* **10**(7), e0131839.

[55] Guillot, M., Moreau, M., Heit, M., Martel-Pelletier, J., Pelletier, J. P. and Troncy, E. [2013], 'Characterization of osteoarthritis in cats and meloxicam efficacy using objective chronic pain evaluation tools', *Veterinary Journal* **196**(3), 360–367.

[56] Hammerla, N., Kirkham, R., Andras, P. and Plötz, T. [2013], On preserving statistical characteristics of accelerometry data using their empirical cumulative distribution, *in* '2013 ACM International Symposium on Wearable Computers', ACM Press, New York, USA, p. 65.

[57] Hammerla, N. Y. and Plötz, T. [2015], 'Let's (not) stick together', *2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* pp. 1041–1051.

[58] Hansen, C. F., Hales, J., Weber, P. M., Edwards, S. A. and Moustsen, V. A. [2017], 'Confinement of sows 24 h before expected farrowing affects the performance of nest building behaviours but not progress of parturition', *Applied Animal Behaviour Science* **188**, 1–8.

[59] Hart, B. L. [1988], 'The behavior of sick animals.', *The Veterinary clinics of North America. Small animal practice* **21**(2), 225–237.

[60] Hart, B. L. [1990], 'Behavioral adaptations to pathogens and parasites: Five strategies', *Neuroscience and Biobehavioral Reviews* **14**(3), 273–294.

[61] Hemsworth, P. H., Coleman, G. J., Barnett, J. L. and Borg, S. [2000], 'Relationships between human-animal interactions and productivity of commercial dairy cows', *Journal of Animal Science* **78**(11), 2821–2831.

[62] Hinchcliff, K. and Geor, R. [2008], 'The horse as an athlete: a physiological overview', *Equine Exercise Physiology: the Science of Exercise in the Athletic Horse* .

[63] Hoey, J., Ploetz, T., Jackson, D., Monk, A. F., Pham, C. and Olivier, P. [2011], 'Rapid specification and automated generation of prompting systems to assist people with dementia', *Pervasive and Mobile Computing (PMC)* **7**(3), 299–318.

[64] Jackson, M., Zeagler, C., Valentin, G., Martin, A., Martin, V., Delawalla, A., Blount, W., Eiring, S., Hollis, R., Kshirsagar, Y. and Starner, T. [2013], FIDO – Facilitating Interactions for Dogs with Occupations: Wearable Dog-Activated Interfaces , *in* 'Proc. ISWC'.

[65] Jarvis, S., Reed, B. T., Lawrence, A. B., Calvert, S. K. and Stevenson, J. [2004], 'Peri-natal environmental effects on maternal behaviour, pituitary and adrenal activation, and the progress of parturition in the primiparous sow', *Animal Welfare* **13**(2), 171–181.

[66] Johnson, R. W. and von Borell, E. [1994], 'Lipopolysaccharide-induced sickness behavior in pigs is inhibited by pretreatment with indomethacin.', *Journal of animal science* **72**(2), 309–314.

[67] Jukan, A., Masip-Bruin, X. and Amla, N. [2016], 'Smart Computing and Sensing Technologies for Animal Welfare: A Systematic Review', *ACM Computing Surveys* **50**(1), 1–27.

[68] Kashiha, M. A., Bahr, C., Ott, S., Moons, C. P. H., Niewold, T. A., Ödberg, F. O. and Berckmans, D. [2013], 'Automatic identification of marked pigs in a pen using image pattern recognition', *Computers and electronics in agriculture* **93**, 111–120.

[69] Keegan, K. G., Arafat, S., Skubic, M., Wilson, D. A. and Kramer, J. [2003], 'Detection of lameness and determination of the affected forelimb in horses by use of continuous wavelet transformation and neural network classification of kinematic data', *American Journal of Veterinary Research* **64**(11), 1376–81.

[70] Keegan, K. G., Kramer, J., Yonezawa, Y., Maki, H., Pai, P. F., Dent, E. V., Kellerman, T. E., Wilson, D. A. and Reed, S. K. [2011], 'Assessment of repeatability of a wireless, inertial sensor–based lameness evaluation system for horses', *American journal of veterinary research* **72**(9), 1156–1163.

[71] Keegan, K. G., MacAllister, C. G., Wilson, D. A., Gedon, C. A., Kramer, J., Yonezawa, Y., Maki, H. and Frank Pai, P. [2012], 'Comparison of an inertial sensor system with a stationary force plate for evaluation of horses with bilateral forelimb lameness', *American Journal of Veterinary Research* **73**(3), 368–374.

[72] Keegan, K. G., Wilson, D. A., Wilson, D. J., Smith, B., Gaughan, E. M., Pleasant, R. S., Lillich, J. D., Kramer, J., Howard, R. D., Bacon-Miller, C., Davis, E. G., May, K. A., Cheramie, H. S., Valentino, W. L. and Van Harreveld, P. D. [1998], 'Evaluation of mild lameness in horses trotting on a treadmill by clinicians and interns or residents and correlation of their assessments with kinematic gait analysis', *American Journal of Veterinary Research* **59**(11), 1370–1377.

[73] Keegan, K. G., Yonezawa, Y., Pai, F., Wilson, D. A. and Kramer, J. [2004], 'Evaluation of a sensor-based system of motion analysis for detection and quantification of forelimb and hind limb lameness in horses', *American Journal of Veterinary Research* **65**(5), 665–670.

[74] Kelmer, G., Keegen, K. G., Kramer, J., Wilson, D. A., Pai, F. P. and Singh, P. [2005], 'Computer-assisted kinematic evaluation of induced compensatory movements resembling lameness in horses trotting on a treadmill', *American Journal of Veterinary Research* **66**(4), 646–655.

[75] Khan, A., Hammerla, N., Mellor, S. and Plötz, T. [2016], 'Optimising sampling rates for accelerometer-based human activity recognition', *Pattern Recognition Letters* **73**, 33–40.

[76] Khan, A., Mellor, S., Berlin, E., Thompson, R., McNaney, R., Olivier, P. and Plötz, T. [2015], Beyond activity recognition, *in* '2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing', number October 2016, ACM, New York, USA, pp. 1155–1166.

[77] Khan, A., Nicholson, J., Mellor, S., Jackson, D., Ladha, K., Ladha, C., Hand, J., Clarke, J., Olivier, P. and Plötz, T. [2014], Occupancy monitoring using environmental context sensors and a hierarchical analysis framework, *in* 'Proc. BuildSys', ACM, New York, NY, USA.

[78] Kim, D., Hilliges, O., Izadi, S., Butler, A. D., Chen, J., Oikonomidis, I. and Olivier, P. [2012], Digits: Freehand 3d interactions anywhere using a wrist-worn gloveless sensor, *in* 'Proceedings of the 25^{th} Annual ACM Symposium on User Interface Software and Technology', UIST '12, ACM, New York, NY, USA, pp. 167–176.

[79] Klein, T. W., DeFries, J. C. and Finkbeiner, C. T. [1973], 'Heritability and genetic correlation: Standard errors of estimates and sample size', *Behavior Genetics* **3**(4), 355–364.

[80] Korhonen, T. [1991], 'Three-dimensional Hall effect accelerometer for recording head movements of freely moving laboratory animals', *Physiology and Behavior* **49**(3), 651–652.

[81] Kyriazakis, I. and Tolkamp, B. J. [2010], Disease, *in* 'The Encyclopedia of Applied Animal Behaviour and Welfare', pp. 16–177.

[82] Ladha, C., Hammerla, N., Hughes, E., Olivier, P. and Ploetz, T. [2013], Dog's life, *in* '2013 ACM international joint conference on Pervasive and ubiquitous computing', ACM, New York, USA, p. 415.

[83] Ladha, C., Hammerla, N. Y., Olivier, P. and Plötz, T. [2013], ClimbAX, *in* '2013 ACM international joint conference on Pervasive and ubiquitous computing', ACM, New York, USA, p. 235.

[84] Laflamme, D. [1997], 'Development and validation of a body condition score system for dogs: a clinical tool', *Canine Practice* **22 (3)**, 10–15.

[85] Larson, S. J. and Dunn, A. J. [2001], 'Behavioral effects of cytokines', *Brain, Behavior, and Immunity* **15**(4), 371–387.

[86] Lascelles, B. D. X., Hansen, B. D., Roe, S., DePuy, V., Thomson, A., Pierce, C. C., Smith, E. S. and Rowinski, E. [2007], 'Evaluation of client-specific outcome measures and activity monitoring to measure pain relief in cats with osteoarthritis', *Journal of Veterinary Internal Medicine* **21**(3), 410–416.

[87] Lawrence, A. B., Petherick, J. C., McLean, K. A., Deans, L. A., Chirnside, J., Gaughan, A., Clutton, E. and Terlouw, E. M. [1994], 'The effect of environment on behaviour, plasma cortisol and prolactin in parturient sows', *Applied Animal Behaviour Science* **39**(3-4), 313–330.

[88] Lawson, S. E. and Marlin, D. J. [2010], 'Preliminary report into the function of the shoulder using a novel imaging and motion capture approach', *Equine Veterinary Journal* **42**(SUPPL. 38), 552–555.

[89] Lee, H.-j., Roberts, S. J., Drake, K. A. and Dawkins, M. S. [2011], 'Prediction of feather damage in laying hens using optical flows and Markov models.', *Journal of the Royal Society, Interface* **8**(57), 489–499.

[90] Leschnik, M., Weikel, J., Möstl, K., Revilla-Fernández, S., Wodak, E., Bagó, Z., Vanek, E., Benetka, V., Hess, M. and Thalhammer, J. G. [2007], 'Subclinical infection with avian influenza A (H5N1) virus in cats', *Emerging Infectious Diseases* **13**(2), 243–247.

[91] Mainau, E., Dalmau, A., Ruiz-de-la Torre, J. L. and Manteca, X. [2009], 'Validation of an automatic system to detect position changes in puerperal sows', *Applied Animal Behaviour Science* **121**(2), 96–102.

[92] Mancini, C., van der Linden, J., Bryan, J. and Stuart, A. [2012], Exploring Interspecies Sensemaking: Dog Tracking Semiotics and Multispecies Ethnography, *in* 'Proc. UbiComp'.

[93] Mancini, C., van der Linden, J., Kortuem, G., Dewsbury, G., Mills, D. and Boyden, P. [2014], UbiComp for Animal Welfare: Envisioning Smart Environments for Kenneled Dogs , *in* 'Proc. UbiComp'.

[94] Marchant, J. N. and Broom, D. M. [1996], 'Factors affecting posture-changing in loose-housed and confined gestating sows', *Animal Science* **63**(3), 477–485.

[95] Marchant, J. N., Broom, D. M. and Corning, S. [2001], 'The influence of sow behaviour on piglet mortality due to crushing in an open farrowing system', *Animal Science* **72**(1), 19–28.

[96] Marchant, J. N., Rudd, A. R., Mendl, M. T., Broom, D. M., Meredith, M. J., Corning, S. and Simmins, P. H. [2000], 'Timing and causes of piglet mortality in alternative and conventional farrowing systems', *Veterinary Record* **147**(8), 209–214.

[97] Marchioro, G. F., Cornou, C., Kristensen, A. R. and Madsen, J. [2011], 'Sows' activity classification device using acceleration data - A resource constrained approach', *Computers and Electronics in Agriculture* **77**(1), 110–117.

[98] Marcroft, C., Khan, A., Embleton, N., Trenell, M. and Ploetz, T. [2015], 'Movement recognition technology as a method of assessing spontaneous general movements in high risk infants', *Frontiers in Neurology* **5**(284).

[99] Mason, G. J. [1991], 'Stereotypies: a critical review', *Animal Behaviour* **41**(6), 1015–1037.

[100] Matheson, S., Thompson, R., Walling, G., Plötz, T., Kyriazakis, I. and Edwards, S. [2016], Relationship between sow conformation, farrowing floor type and posture change characteristics using accelerometer data, *in* '50$^{\text{th}}$ Annual Meeting of the International Society for Applied Ethology'.

[101] Matsuura, A., Irimajiri, M., Matsuzaki, K., Hiraguri, Y., Nakanowatari, T., Yamazaki, A. and Hodate, K. [2013], 'Method for estimating maximum permissible load weight for Japanese native horses using accelerometer-based gait analysis', *Animal Science Journal* **84**(1), 75–81.

[102] Matthews, S. G., Miller, A. L., Clapp, J., Plötz, T. and Kyriazakis, I. [2016], 'Early detection of health and welfare compromises through automated detection of behavioural changes in pigs', *Veterinary Journal* **217**, 43–51.

[103] Matthews, S. G., Miller, A. L., Plötz, T. and Kyriazakis, I. [2017], 'Automated tracking to measure behavioural changes in pigs for health and welfare monitoring', *Scientific Reports* **7**(1), 17582.

[104] Mendl, M., Burman, O. H. P. and Paul, E. S. [2010], 'An integrative and functional framework for the study of animal emotion and mood', *Royal Society B: Biological Sciences* **277**(1696), 2895–2904.

[105] Moller, A., Roalter, L., Diewald, S., Scherr, J., Kranz, M., Hammerla, N., Olivier, P. and Plotz, T. [2012], GymSkill: A personal trainer for physical exercises, *in* '2012 IEEE International Conference on Pervasive Computing and Communications', IEEE, pp. 213–220.

[106] Moorman, V. J., Reiser, R. F., Peterson, M. L. ., Mcllwraith, C. W. and Kawcak, C. E. [2013*a*], 'Effect of forelimb lameness on hoof kinematics of horses at a trot', *American Journal of Veterinary Research* **74**(9), 1183–91.

[107] Moorman, V. J., Reiser, R. F., Peterson, M. L. ., Mcllwraith, C. W. and Kawcak, C. E. [2013*b*], 'Effect of forelimb lameness on hoof kinematics of horses at a walk', *American Journal of Veterinary Research* **74**(9), 1192–7.

[108] Mundl, W. J. [1966], 'Activity of small animals measured with accelerometer', *Medical & Biological Engineering* **4**(2), 209–212.

[109] Mundl, W. J. and Malmo, H. P. [1979], 'An accelerometer for recording head movement of laboratory animals', *Physiology and Behavior* **23**(2), 391–393.

[110] Niggli, W. and Holstein, G. [2003], *Dressage: A Guideline for Riders and Judges*, Marlborough: The Crowood Press.

[111] Oczak, M., Maschat, K., Berckmans, D., Vranken, E. and Baumgartner, J. [2015], 'Classification of nest-building behaviour in non-crated farrowing sows on the basis of accelerometer data', *Biosystems Engineering* **140**, 48–58.

[112] Oczak, M., Viazzi, S., Ismayilova, G., Sonoda, L. T., Roulston, N., Fels, M., Bahr, C., Hartung, J., Guarino, M., Berckmans, D. and Vranken, E. [2014], 'Classification of aggressive behaviour in pigs by activity index and multilayer feed forward neural network', *Biosystems Engineering* **119**, 89–97.

[113] Oliviero, C., Pastell, M., Heinonen, M., Heikkonen, J., Valros, A., Ahokas, J., Vainio, O. and Peltoniemi, O. A. T. [2008], 'Using movement sensors to detect the onset of farrowing', *Biosystems Engineering* **100**(2), 281–285.

[114] Pastell, M., Hietaoja, J., Yun, J., Tiusanen, J. and Valros, A. [2013], Predicting farrowing based on accelerometer data, PhD thesis, Ed.: D. Berckmans and J. Vandermeulen.) KU, Leuven (ISBN: 978-90-8826-333-0), 243–252.

[115] Pastell, M., Hietaoja, J., Yun, J., Tiusanen, J. and Valros, A. [2016], 'Predicting farrowing of sows housed in crates and pens using accelerometers and CUSUM charts', *Computers and Electronics in Agriculture* **127**, 197–203.

[116] Pastell, M., Tiusanen, J., Hakojärvi, M. and Hänninen, L. [2009], 'A wireless accelerometer system with wavelet analysis for assessing lameness in cattle', *Biosystems Engineering* **104**(4), 545–551.

[117] Peeters, M. E. [2013], 'Feline Hyperthyroidism', *Small Animal Soft Tissue Surgery* **31**(5), 17–26.

[118] Pfau, T., Ferrari, M., Parsons, K. and Wilson, A. [2008], 'A hidden Markov model-based stride segmentation technique applied to equine inertial sensor trunk movement data', *Journal of Biomechanics* **41**(1), 216–220.

[119] Pitts, A. D., Weary, D. M., Fraser, D., Pajor, E. A. and Kramer, D. L. [2002], 'Alternative housing for sows and litters Part 5. Individual differences in the maternal behaviour of sows', *Applied Animal Behaviour Science* **76**(4), 291–306.

[120] Ploetz, T., Moynihan, P., Pham, C. and Olivier, P. [2010], Activity Recognition and Healthier Food Preparation, *in* 'Activity Recognition in Pervasive Intelligent Environments', Atlantis Press.

[121] Plötz, T., Hammerla, N. Y., Rozga, A., Reavis, A., Call, N. and Abowd, G. D. [2012], Automatic assessment of problem behavior in individuals with developmental disabilities, *in* '2012 ACM Conference on Ubiquitous Computing', ACM Press, New York, USA, p. 391.

[122] Powers, D. M. [2011], 'Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation'.

[123] Ramos, A. and Mormède, P. [1997], 'Stress and emotionality: A multidimensional and genetic approach', *Neuroscience and Biobehavioral Reviews* **22**(1), 33–57.

[124] Ringgenberg, N., Bergeron, R. and Devillers, N. [2010], 'Validation of accelerometers to automatically record sow postures and stepping behaviour', *Applied Animal Behaviour Science* **128**(1-4), 37–44.

[125] Rodenburg, T. B., Bijma, P., Ellen, E. D., Bergsma, R., De Vries, S., Bolhuis, J. E., Kemp, B. and Van Arendonk, J. A. M. [2010], 'Breeding amiable animals? Improving farm animal welfare by including social effects in breeding programmes', *Animal Welfare* **19**(SUPPL. 1), 77–82.

[126] Ruiz-Garcia, L., Lunadei, L., Barreiro, P. and Robla, I. [2009], 'A Review of Wireless Sensor Technologies and Applications in Agriculture and Food Industry: State of the Art and Current Trends', *Sensors* **9**(6), 4728–4750.

[127] Rushen, J., Pombourcq, E. and de Passillé, A. M. [2007], 'Validation of two measures of lameness in dairy cows', *Applied Animal Behaviour Science* **106**(1-3), 173–177.

[128] Sellen, A. J., Fogg, A., Aitken, M., Hodges, S., Rother, C. and Wood, K. [2007], Do life-logging technologies support memory for the past?, *in* 'SIGCHI conference on Human factors in computing systems', ACM Press, New York, USA, p. 81.

[129] Serrano-Cinca, C., Fuertes-Callén, Y. and Mar-Molinero, C. [2005], *Measuring DEA efficiency in Internet companies*, Vol. 38, Academic Press.

[130] Špinka, M., Illmann, G., De Jonge, F., Andersson, M., Schuurman, T. and Jensen, P. [2000], 'Dimensions of maternal behaviour characteristics in domestic and wildxdomestic crossbred sows', *Applied Animal Behaviour Science* **70**(2), 99–114.

[131] Stamp Dawkins, M. [2007], *Observing Animal Behaviour*, Oxford: Oxford University Press.

[132] Statham, P., Green, L., Bichard, M. and Mendl, M. [2009], 'Predicting tail-biting from behaviour of pigs prior to outbreaks', *Applied Animal Behaviour Science* **121**(3-4), 157–164.

[133] Sturgess, K. [2007], Disease And Welfare, *in* I. Rochlitz, ed., 'The welfare of cats', Dordrecht: Springer, pp. 205–225.

[134] Sundermeyer, M., Schlüter, R. and Ney, H. [2012], Lstm neural networks for language modeling, *in* 'Thirteenth Annual Conference of the International Speech Communication Association'.

[135] Szyszka, O., Tolkamp, B. J., Edwards, S. A. and Kyriazakis, I. [2013], 'The effects of acute versus chronic health challenges on the behavior of beef cattle 1', *Journal of animal science* **90**(12), 4308–4318.

[136] Thompson, R., Matheson, S. M., Plötz, T., Edwards, S. A. and Kyriazakis, I. [2016], 'Porcine lie detectors: Automatic quantification of posture state and transitions in sows using inertial sensors', *Computers and Electronics in Agriculture* **127**, 521–530.

[137] Tinsley, H. E. and Weiss, D. J. [1975], 'Interrater reliability and agreement of subjective judgments', *Journal of Counseling Psychology* **22**(4), 358–376.

[138] Tolkamp, B. J., Allcroft, D. J., Austin, E. J., Nielsen, B. L. and Kyriazakis, I. [1998], 'Satiety splits feeding behaviour into bouts', *Journal of Theoretical Biology* **194**(2), 235–250.

[139] Ursinus, W. W., Van Reenen, C. G., Kemp, B. and Bolhuis, J. E. [2014], 'Tail biting behaviour and tail damage in pigs and the relationship with general behaviour: Predicting the inevitable?', *Applied Animal Behaviour Science* **156**, 22–36.

[140] Valentin, G., Alcaidinho, J., Howard, A., Jackson, M. M. and Starner, T. [2015], 'Towards a Canine-Human Communication System Based on Head Gestures', *Second International Congress on Animal-Computer Interaction* .

[141] Valentin, G., Freil, L., Alcaidinho, J., Zuerndorfer, J., Mason, C. and Jackson, M. M. [2015], 'Wearable Alert System for Mobility-Assistance Service Dogs', *Journal on Technology and Persons with Disabilities* pp. 184–203.

[142] Van Der Linden, J., Schoonderwaldt, E., Bird, J. and Johnson, R. [2011], 'MusicJacket - Combining motion capture and vibrotactile feedback to teach violin bowing', *IEEE Transactions on Instrumentation and Measurement* **60**(1), 104–113.

[143] Viazzi, S., Bahr, C., Van Hertem, T., Schlageter-Tello, A., Romanini, C. E. B., Halachmi, I., Lokhorst, C. and Berckmans, D. [2014], 'Comparison of a three-dimensional and two-dimensional camera system for automated measurement of back posture in dairy cows', *Computers and Electronics in Agriculture* **100**, 139–147.

[144] Vicon [n.d.], '3d motion capturing', online (02-2015): `www.vicon.com`.

[145] Wakeling, J., Moore, K., Elliott, J. and Syme, H. [2008], 'Diagnosis of hyperthyroidism in cats with mild chronic kidney disease', *Journal of Small Animal Practice* **49**(6), 287–294.

[146] Wang, J. S., Huang, Y. S., Wu, M. C., Lai, Y. Y., Chang, H. L. and Young, M. S. [2005], 'Quantification of pre-parturition restlessness in crated sows using ultrasonic measurement', *Asian-Australasian Journal of Animal Sciences* **18**(6), 780–786.

[147] Wang, Z., Mirbozorgi, S. A. and Ghovanloo, M. [2015], Towards a kinect-based behavior recognition and analysis system for small animals, *in* '2015 IEEE Biomedical Circuits and Systems Conference', IEEE, pp. 1–4.

[148] Watanabe, N., Sakanoue, S., Kawamura, K. and Kozakai, T. [2008], 'Development of an automatic classification system for eating, ruminating and resting behavior of cattle using an accelerometer', *Grassland Science* **54**(4), 231–237.

[149] Watanabe, S., Izawa, M., Kato, A., Ropert-Coudert, Y. and Naito, Y. [2005], 'A new technique for monitoring the detailed behaviour of terrestrial animals: A case study with the domestic cat', *Applied Animal Behaviour Science* **94**(1-2), 117–131.

[150] Weary, D. M., Huzzey, J. M. and Von Keyserlingk, M. A. G. [2009], 'Board-invited Review: Using behavior to predict and identify ill health in animals', *Journal of Animal Science* **87**(2), 770–777.

[151] Wechsler, B. and Hegglin, D. [1997], 'Individual differences in the behaviour of sows at the nest-site and the crushing of piglets', *Applied Animal Behaviour Science* **51**(1-2), 39–49.

[152] White, K. R., Anderson, D. M. and Bate, L. A. [1996], 'Increasing piglet survival through an improved farrowing management protocol', *Canadian Journal of Animal Science* **76**(4), 491–495.

[153] Williams, C. K. I. [2003], *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, Vol. 98, MIT Press.

[154] Witte, T. H. [2004], 'Determination of peak vertical ground reaction force from duty factor in the horse (Equus caballus)', *Journal of Experimental Biology* **207**(21), 3639–3648.

[155] Xenophon [n.d.], Πϵρά¡ű ά¡ůππιχά£Ęς *- peri hippikēs (Engl.: On Horsemanship).*

[156] Yonezawa, K., Takashi, M. and Rekimoto, J. [2009], Cat @ Log : Sensing Device Attachable to Pet Cats for Supporting Human-Pet Interaction, *in* 'International Conference on Advances in Computer Entertainment Technology', ACM Press, New York, USA, pp. 149–156.

[157] Zaleski, H. M. and Hacker, R. R. [1993], 'Variables related to the progress of parturition and probability of stillbirth in swine.', *The Canadian veterinary journal* **34**(2), 109–113.