A Chromatic Approach to Fatigue Risk Quantification in the Workplace

By

Jacob Thomas Merriman

A thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

Department of Electrical Engineering and Electronics

May 2023

Contents

Acknowledgements

I would like to express my heartfelt thanks to my supervisors, Jim Humphries and Roberto Ferrero, for their unwavering guidance, invaluable support, and immense knowledge throughout the entire duration of my journey. Their expertise and insightful feedback have been instrumental in shaping this thesis and my overall growth as a researcher. I am grateful for their constant encouragement, patience, and willingness to dedicate their time to my academic pursuits.

Thank you to the surgical team for participating in my research. Without you, my work would not have been complete. Additionally, my thanks go to Alison Goodyear and the rest of the administration team, for your insightful knowledge and kindness throughout my journey.

My thanks also go to Fatigue Management International (FMI) for their help with the development of the algorithm and also to MYANT for providing the equipment used in the study, both of which made the study possible.

I would like to express my gratitude to my dear friends from Grimsby and beyond who have been an unwavering source of support, encouragement and escapism before, during and after the COVID pandemic. Their presence in my life has made a significant impact, both personally and academically, and I am immensely grateful for their unwavering support. Thank you for always being there to lend an empathetic ear, share in the highs and lows, and provide the much-needed motivation to keep pushing forward. Their friendship has been a constant source of inspiration, and I am grateful for the countless hours we have spent on video calls, sharing laughter, and celebrating milestones together.

Finally, I would like to dedicate a special section of gratitude to my loving partner Shan, my siblings, my parents and my grandparents for their unwavering love, encouragement, and continuous belief in my abilities. Their support and understanding have been the pillars of my strength throughout this challenging journey. To those who are not with us anymore, I think of you every day.

Abstract

Surgeons work around the clock with on-call hours leading to working overtime. Additionally, long hours with constant interruptions, complex patient care needs and draining emotional interactions with patients and families contribute to work-related fatigue. Numerous studies have made countless efforts to mitigate the effects of fatigue on ICU personnel, however, there is very little scientific proof to suggest these conclusions work. There is a lack of non-invasive technology that can monitor fatigue detection in the workplace, without the use of questionnaires.

An advisory system is developed to predict the onset of fatigue in shift workers, namely surgeons, to offer strategic implementations of breaks to mitigate the beginning of potential errors that are made in the workplace. Using chromatic data processing, this thesis describes the development of an algorithm based on physiological and systematic indicators of fatigue.

The physiological factors are derived from an understanding of the predicted Karolinska Sleepiness Scale (KSS) in shift workers through a pilot study, sleep quality the previous night and Heart Rate Variability. The systematic measures consider the effect of time-on-shift. The time-on-shift element may be modified with break opportunities or frequency-domain HRV measures and activity levels, which are shown to reduce the KSS scale in this work.

The algorithm(s) described produce a strong correlation against KSS scores, thus mitigating the need for surgeons to fill out questionnaires on shift. They demonstrate high accuracy at predicting fatigue levels and trigger accurate alarms, alerting the user to take a break when appropriate.

Chapter 1

Introduction

1.1 Research overview

With increased literature, momentous evidence points towards preventable medical errors in surgeries (Khanade and Sasangohar, 2017). Occupational symptoms and injuries are sustained over a surgical career. They are under-reported, yet they impact daily surgical practice (Voss et al., 2016), causing surgeons to take leave of absence and retire earlier than planned (Stucky et al., 2018). Work-related pain is present in up to 87% of surgeons (Park et al., 2010), and burnout is reported in up to 40% (Shanafelt et al., 2009), both implicated as patient safety hazards. Studies in the UK, USA and Australia broadly support around 10% of all patients admitted to hospitals coming to harm due to medical error (Berwick and Leape, 1999; Bethune and Francis, 2015). However, retrospective reviews show that only 6% of all medical errors are related to a lack of knowledge and technical ability, whereas 73% are related to human factors such as team working, decision-making, stress and fatigue (Van Watendonk et al., 2010).

There is a reason why surgeon fatigue is underreported. Medical errors are the third leading cause of death in the US (CNBC, 2018). Yet, even to this date, it is still being determined whether a surgeon is fatigued (or too fatigued) at work. This makes treating,

administering breaks or reducing motor vehicle accidents difficult. Society trusts national healthcare to provide efficient judgement, treatment and monitoring of an individual's health. However, in medical practices, the issue of sleep deprivation has captured the public's attention (Santini et al., 2011).

Intensive Care Unit [ICU] staff, such as surgeons and junior doctors, work around the clock with on-call hours leading to working overtime, limiting the ability to control the number of hours worked, causing work to lead to fatigue and sleep disturbances (Bae et al., 2013). Long hours with constant interruptions, complex patient care needs and draining emotional interactions with patients and families contribute to work-related fatigue for ICU staff (Cockerham et al., 2018). Of note, the COVID-19 pandemic has been shown to increase the prevalence of emotional exhaustion and depersonalisation (Houdmont et al., 2021), prompting urgent measures to address the ongoing issue of surgeon burnout in the years to come.

Surgeons' techniques and knowledge are perfected and refined annually to improve patient outcomes. However, human physiology cannot keep up in a world that is forever advancing. Newer surgical techniques, which include the role of robotics such as Minimally Invasive Surgery (MIS), have great benefits for patients but increase the physical workload for the surgeons (Alleblas et al., 2017). Recent ergonomic studies show that MIS surgeons are significantly more likely to experience musculoskeletal symptoms than those performing open surgery (Stucky et al., 2018). In a delicate profession where circadian rhythms are obsolete and neglected (Knutsson, 2003), advancements in technology also need to be fixated on the surgeons performing the operations. We are now beginning to understand the need for self-care in healthcare professionals (Bethune and Francis, 2015), as many surgeons experience work-related injuries. Some of these require leave of absence, early retirement and even surgery (Stucky et al., 2018).

Fatigue can be derived from physical exertion, mental persistence, sleep deprivation and environmental constructs, inevitably leading to negative consequential

behaviours/actions in a workplace and errors are likely to be catastrophic. Therefore, mediating fatigue in the workplace is of utmost importance, especially when mistakes can have fatal or financial implications. The need for objective methods to diagnose and track fatigue in the workplace is critical when individuals are motivated not to report their amount of sleep or their level of fatigue (Haghighi and Yazdi, 2015). Financial or social pressures may result in false recordings of self-reported data to alleviate pressure on themselves, consequently endangering others, even themselves.

Surgeons, in particular, are usually in control of when they take breaks during their shifts by simply taking a break when they have the time. In instances where their perceived workload is too significant, and they are not taking breaks, an advisory system may be able to alleviate them from responsibilities and offer an opportunity to remove themselves from a potentially hazardous situation where their attention span continues to decline as a result of increased fatigue. Breaks being taken at appropriate points during their shift in the evolution of fatigue may have a substantially more beneficial effect on their current fatigue condition, i.e. surgeons shouldn't take breaks simply when they can fit them in (Bosch and Sonnentag, 2019).

1.2 Current research

There have been a number of proposed strategies to mitigate the effects of fatigue on ICU personnel. These include proposing self-education on the health risks associated with fatigue (Caruso et al., 2017), changes to the NHS to reduce distress and improve patient care (Farquharson et al., 2012), strategic intervention plans to rescue further personal and organisational damage (Jarrad et al., 2018), introduce self-care coping strategies into training modules (Hevezi, 2015) or incorporating rest breaks, naps and exercise into shift patterns (Kaliyperumal et al., 2017). However, there is very little scientific proof to suggest these conclusions work; and therefore organisations and managers are reluctant to change rotas and work patterns to test these interventions.

Due to these implications, alternative avenues must be explored to reduce fatigue, providing clear, objective data that can be used to monitor and potentially mitigate the

effects of fatigue and initial insight into an individual's physiology for various uses, whether the demographic is surgeons, athletes, labour workers or patients. Yet to enter the market is a device which can monitor and potentially mitigate the effects on fatigue of surgeons whilst on-call.

There is a plethora of systematic reviews looking into surgeon fatigue (Sturm et al., 2011; Mansukhani et al., 2012; Harris et al., 2015; Gates et al., 2018; Stucky et al., 2018; Dias et al., 2018) from a variety of aspects including; cognitive workload, ergonomics, sleep deprivation and musculoskeletal injuries. Sturm et al. looked at performance and patient outcomes as a result of fatigue, revealing mixed results. Whilst some studies found no difference in clinical performance with sleep deprivation, others found increases in complications and errors. The presence of multiple potential confounding factors makes assessment of these outcomes difficult.

In recent years, Bio-Mathematical Models (BMMs) have been developed in laboratories to predict fatigue and performance whilst on shift (Dawson et al., 2011). These models predict the general construct of fatigue based on sleep-wake behaviour and working-time arrangements. Yet, the term 'fatigue' is used loosely and relates more to perceptions of tiredness (Dawson et al., 2011). Even though they are used in some occupations, such as piloting, they are simply opportunities for sleep administered by the employer. As a global risk assessment tool, they are appropriate but lack specificity and sensitivity at the individual worker level (Dawson et al., 2017). In laboratories, they are great assessors of performance decrements with practical examinations of driving behaviours or generic task performances. However, they may not be an accurate proxy for the performance of a task in the workplace (Dawson et al., 2014), where naturally occurring stressors may not be so evident (Griffiths et al., 2017). The greatest downfall of BMMs is an organisation's reluctance to adopt such a model despite over 50 years of evolution (Dawson et al., 2011). Minimal evidence is published in the real-world application of BMMs despite being a potentially promising tool for fatigue mitigation with considerable public safety and health implications (James et al., 2018).

Limited studies exist monitoring real-time fatigue in ICU personnel. Johnston et al. (2018) used personal diary assistants (PDAs) to track the nurses' fatigue levels throughout the shift. They concluded that self-perceived levels of reward and control over their shift predicted fatigue accumulation rather than physiological measures, namely energy expenditure. Triangulating physiological and psychological constructs is necessary to alleviate the onset of workplace performance decrements and negative emotions. Another study replicated similar results in field-based research. Griffiths et al. (2017) studied forestry workers whose injury risk is high due to fatigue. The study found that a chest strap can collect meaningful heart rate variability data to identify fatigue at work. However, no literature yet examines the real-time monitoring of surgeon fatigue.

1.3 Research contributions

This thesis focuses on the development of real-time fatigue detection in the workplace. An advisory system is developed to predict the onset of fatigue in shift workers, namely surgeons, to offer strategic implementations of breaks to mitigate the beginning of potential errors that are made in the workplace. The backbone of the algorithm for the advisory system comprises physiological and systematic measures;

Physiological factors:

a) Predicted Karolinska Sleepiness Scale (KSS) in shift workers, which varies depending on the shift type (day, night or 24-hour),
b) Sleep quality from the previous night,
c) Biomarkers that make the algorithm sensitive to the individual, unlike BMM's.

Systematic factors:

a) Time on Shift, which rises exponentially (Koh et al., 2007; Johnston et al., 2018) and taking breaks for rest opportunities.

As there is no absolute measure for sleepiness, subjective fatigue assessment scales such as the KSS, Stanford sleepiness scale (SSS) or other Likert scales are accepted as reasonably accurate indicators of fatigue/sleepiness (Johns, 2010). Physiological

factors can depict an individual's stress or fatigue on their shift. Biomarkers such as Heart Rate (HR), Heart Rate Variability (HRV) or Skin Temperature (ST) are analysed and compared against the KSS for their correlation against the 9-point Likert scale, which is used to detect sleepiness in the workplace. Other considerations include sleep quality from the previous night before working and the predicted circadian rhythm, respective to shift type. The systematic measures consider the effect of time on shift but, the time on shift element may be modified with break opportunities or frequency-domain HRV measures and activity levels, which are shown to reduce the KSS scale in this work. Multiple cross referenced models are applied, to increase the accuracy of fatigue onset, whilst considering all possible non-invasive contributors to fatigue.

1.4 Objectives and benefits

There are no absolute measures of workplace fatigue (Haghighi and Yazdi, 2015), a construct deriving from cognitive and physiological energy depletion. However, biomarkers of fatigue and subjective well-being can be monitored around the clock to predict the onset of fatigue over time to suggest when to take a break from work at the most opportune moment. This seems to be the most appropriately fitting solution with the current technology and resources we have to date. One of the major considerations in the quantification of workplace fatigue is associated with the invasiveness of technology administered during the shift. It is necessary therefore, to minimise the application of invasive technology and work schedule deviation in order to predict the onset of fatigue in real-time.

The aims of thesis are as follows:

1. Identify the contributors to fatigue within the workplace and possible mitigation strategies.
2. Produce an advisory system to predict the onset of fatigue in the workplace.
3. Utilise the engineering technique known as 'chromatic processing' for the building blocks of the algorithm.

This research aims to develop an advisory system which can predict the onset of fatigue in the workplace in an unobtrusive manner that does not interfere with the user's work commitments. The users in this research specifically being surgeons. For the administration of this fatigue risk management system, key considerations must be adhered to:-

1. Comfortable user experience
2. Economical cost
3. Unobtrusive to the user's work
4. Sensitive to the individual and shift type
5. Real-time detection of fatigue onset
6. Act as an advisory system for strategic breaks
7. Ease of adoption by both user and employer

No technology fulfilling the above considerations is currently available to monitor fatigue for shift workers in the workplace. The physiological and systematic systems produce a series of complex data outputs. The data needs to be processed and represented in such a way to be easily interpreted by the user in a mathematically economical way so that it preserves the battery life of the wearable device, so that it can be worn in excess of 24 hours. This thesis proposes an advisory system to be used in the workplace using a novel engineering technique, for a surgical cohort.

A novel method of analysing complex signals known as 'Chromatic analysis' has been developed at University of Liverpool. The method is utilised in the signal processing of complex data sets from biometric sensors via three triangular filters, as well as modified truncated filters, where the signal processing technique transforms complex data sets into the parameters *H*, *L* and *S*. The transformation of the data sets allows visual representation of the data in polar plots or line graphs and predicts workplace fatigue through individually weighted variables within defined parameters. *H* represents the dominant or average frequency within the signal spectrum, *L* indicates the overall strength or bandwidth of the signals, and *S* represents the spread of the signals (Zhang et al., 2004). The proposed solutions utilise the H and L parameters along with other factors in the timely diagnosis of fatigue.

The novelty of this thesis is as follows:

1. Identification of fatigue contributors and mitigation strategies within the workplace in surgeons.
2. Real time monitoring of workplace fatigue using Linear Mixed Model approach and Chromatic processing.
3. Production of advisory system which can trigger multiple alarms during one's shift for the timely mitigation of workplace fatigue.

1.5 Thesis outline

The thesis is split into eight chapters. Chapter 2 outlines the previous literature on the topic, which outlines the prevalence of workplace fatigue in the selected cohort during work and identifies the circadian rhythms utilised in the advisory system. The chapter also describes the chromatic analysis method. Chapter 3 explains preliminary research conducted by the research team to outline the underlying causes of fatigue in the workplace, as well as a pilot study. Chapter 4 discusses the methodology used for the advisory system. Chapter 5 displays experimental data collection on the surgeons. Chapter 6 discusses results using the chromatic processing technique, including the concepts of the algorithm and examples of its execution in the workplace. Chapter 7 includes demonstration of the algorithm, including when the algorithm would (or would not) signal breaks for the individuals. Finally, Chapter 8 summarises the work, research outcomes, and future recommendations to strengthen this work further.

Chapter 2

Literature review of previous work

2.1 Introduction

This chapter outlines the previous research conducted in relation to surgeon fatigue, considering objective, physiological and subjective methods to identify fatigue in the workplace, for example heart rate [HR] (Holdsworth and Evens, 2017), circadian rhythm (Haghighi and Yazdi, 2015) and the Karolinska Sleepiness Scale [KSS] (Shahid et al., 2011), respectively. The difficulty in tracking fatigue in the workplace is described, as well as an efficient information extraction approach based upon chromaticity for identifying fatigue. Chromaticity and its application are outlined and reviewed in this section. All information relating to statistical tests used in this chapter can be found in Appendix A.

2.2 Surgeon work related fatigue

Despite extensive literature investigating musculoskeletal fatigue of surgeons, there are other aspects of a surgical career which can lead to fatigue which are seldom reported, such as mental exhaustion, sleep deprivation or compassion fatigue. The need for objective methods to diagnose and/or track workplace fatigue in the workplace is especially important in circumstances when individuals are motivated not to report truthfully either their amount of sleep or their level of fatigue in the workplace (Haghighi and Yazdi, 2015).

Due to these aforementioned implications, alternative avenues must be explored to reduce fatigue, those of which provide objective, clear data sets. Wearable technology is on the rise, offering substantial insight into an individual's physiology for a variety of uses, whether the demographic be athletes, labour workers or patients. The gold standard for measuring fatigue uses brainwave activity (Electroencephalography), however this is extremely difficult to monitor in a real-world application. Many gold-standards are invasive and unable to be implemented in field-based research, such as the 12-lead ECG for the monitoring of heart rate. Additionally, initial conversation with our surgical liaison explained that NHS (UK) surgeons are unable to wear anything below the elbows or above the neck, which limits the apparatus (e.g. watches) used for collection of data in real time. Wearable technology can provide thorough insight into one's physiology and can identify the onset of fatigue for anybody.

2.2.1 Monitoring strategies for workplace fatigue

Two current methods to monitor fatigue are from an objective or subjective view. Objective measures analyse real time biomarkers on the body, such as heart rate, core body temperature or reaction time and subjective measures include questionnaires such as the Karolinska Sleepiness Scale. One journal (Anwer et al., 2021) highlights the complex relationship between physical and mental fatigue and identified the importance of these metrics being used in combination, more accurately identifying fatigue onset. A previous study (Johnston et al., 2018) cross-referenced subjective and objective fatigue markers in nurses and found that reward was negatively correlated with fatigue and motivational situations are less likely to induce fatigue, whereas demanding situations can exacerbate fatigue. Additionally, Hockey (2011) highlights that high levels of control in the workplace are also negatively correlated with fatigue, regardless of whether effort is high as this generally increases alertness and does not necessarily mean there are reductions in work efficiency or cognitive functionality.

2.2.1.1 Objective markers

Heart rate (HR), heart rate variability (HRV) (Pichot et al., 2002) and skin temperature (ST) (Haghighi and Yazdi, 2015) have previously been used successfully to identify fatigue in the workplace. Johnston et al. (2018) provided the most comprehensive

examination of subjective fatigue in the workplace with nurses, using a heart rate monitor and questionnaires.

HR data captured throughout the course of a work period can be used for indication of task demand (Griffiths et al., 2017). Higher HR typically mirrors high periods of physical activity due to an increased demand for oxygenated blood to the muscles. Interpretation of HR at work can individualise the data and determine periods of high and low activity from deviation at rest. A HR which goes below 60 beats per minute (BPM) at rest is typical of an individual with good cardiovascular fitness. These instances of low HR may coincide with lack of stress due to the relaxed nature of the individual during a heavily demanding job. Resting HR over 90BPM can be an indicator of extreme stress or other physiological conditions which may require consultation, and is outside the scope of this research. High levels of demanding work with lots of physical movement will inevitably increase HR due to the muscles needs for oxygen, however there may be instances of work which indicate low levels of movement but a high HR. These instances may occur as a response to a stressful situation, such as performing surgery or having difficult conversations with family or friends of patients. HR is calculated from the time in between beats, also known as the 'R-R Interval' (RRI). HR can be a valuable tool for identifying periods of high intensity in the workplace, however HR alone is not as insightful to identify periods of high stress. Additional measures using HR are available but also mandatory in order to obtain a clearer picture of one's subjective fatigue levels.

A recent overview of metrics and normative values supports the use of HRV for objective assessment of psychological health and stress (Shaffer and Ginsberg, 2017). HRV indicates the ability of the heart to react to environmental and physiological stimuli through measuring the fluctuation between heart beats in milliseconds [ms]. Low HRV is associated with an impaired regulatory and homeostatic autonomic nervous system [ANS] which reduces the body's ability to cope with internal and external stressors. Holdsworth and Evens (2017) conducted a systematic review on HRV as a useful marker of stress and fatigue in ICUs. The study found that with high stress and burnout rates and low job control that HRV is a useful marker of stress and

fatigue in these settings, thereby helping to optimise clinician performance and the standard of healthcare.

HRV may be represented in either the time or the frequency domain. Time-domain indices of HRV quantify variability in measurements of the RRI. The RRI (R-R interval) measures the time elapsed between two successive 'R' peaks on the QRS signal of an electrocardiogram (detailed in Figure 4.3). The root mean square of successive differences (RMSSD) has been used extensively for short-term ECG analysis in the time domain. RMSSD reflects the RRI variance in HR. Lower values of RMSSD indicate higher levels of sympathetic nervous system (SNS) activation. The SNS prepares the body for physical activity, through redirection of oxygenated blood around the body in response to physical or mental stimuli. Frequency-domain measures estimate the distribution of relative power into frequency bands, ranging from ultra-low-frequency ($\leq$0.003Hz) to high frequency (0.15 – 0.4Hz). Power is the signal energy found within frequency bands. The common method for identification of HRV for frequency-domain measures is the Low-frequency/High-frequency Ratio (LF/HF Ratio). The LF component contributes to the effects of both the parasympathetic nervous system (PNS) and the sympathetic nervous system (SNS). The HF component exclusively contributes to the PNS. A high ratio reflects the SNS dominance and withdrawal of the PNS, resulting in increased HR and onset of stress.

The psychomotor vigilance test (PVT) is a sustained-attention, reaction-timed task that measures the speed (ms) that subjects respond to a visual stimulus. Lerman et al. (2017) implemented the PVT in the workplace of medical staff and found that those who were deemed as 'fit for work' (7 or fewer errors) had significantly fewer PVT errors than employees determined to be 'unfit' (12+ errors) based off previous studies findings (Van Dongen et al., 2003; Basner et al., 2011). Night shift workers have been shown to have slower response times and lower accuracy levels than day shift workers due to circadian misalignment (Niu et al., 2011) and whilst this does not necessarily mean more errors will occur during the shift, it increases the likelihood of pervasive medical errors.

Another objective marker is skin temperature. Heat exhaustion is a result of prolonged exposure to high temperatures, particularly combined with high humidity or strenuous exercise (Mayo Clinic, 2023) which results in heavy sweating and the body overheating. Without treatment, heat exhaustion can be life threatening (Mayo Clinic, 2023). However, for shift workers in the UK the risks of heat exhaustion are minimal. Skin Temperature (ST) has been shown to decrease slightly in participants who are sleep deprived (Babkoff et al., 1991, Miro et al., 2002), which may be due to thermoregulatory disruption and decreased metabolism (Horne, 1988). These results were observed in sleep deprived participants over a 48 hour period and as such, will not be observed in the current study. ST is difficult to measure and seldom reported in field-based research, due to environmental factors such as weather or clothing, which may affect temperature readings due to convection or conduction of the sensor depending on the position of the sensor. Nonetheless, ST will be measured in this study for its potential impact of workplace fatigue in field-based research.

Whilst previous research has monitored temperature as a parameter of fatigue (Miro et al., 2002), changes in temperature are generally associated with physical fatigue, and less so an indicator of stress (Aryal and Ghahramani, 2017). On the other hand, ST has been shown to increase after a Stroop task when inducing mental fatigue, compared with a control group (Roelands et al., 2016).

2.2.1.2 Subjective markers

Whilst biomarkers could singlehandedly predict onset of fatigue, results may be falsely perceived due to the inevitability of time awake (Johnston et al., 2018). Identifying workplace fatigue as opposed to physiological fatigue requires an individual's subjective feelings in order to determine motivational control of a situation and thus reluctance to work (Hockey, 2011; Johnston et al., 2018). Whilst there are potential biomarkers to monitor cognitive function such as monoamines, the final concentrations extracted from biological fluids are not in equilibrium with monoamines in the brain due to the blood-brain barrier, as demonstrated by over 100 citations (Hinz et al., 2011). This means that the perceived levels of monoamines will not correlate with the levels

in the brain. Therefore, compliance with the questions based on an individual's feelings in the workplace at present seems the most viable way to obtain cognitive fatigue (Johnston et al., 2018).

Johnston et al. (2018) based their questions given to the nurses through manipulation of previous literature; Hockey and Earle (2006) found that high levels of control can reduce fatigue in highly demanding work tasks whilst assessing 'workload' and 'control'. Hulsheger (2016) extends the recover theory and found that trajectories of daily fatigue are influenced by sleep quality from the previous night whilst assessing 'fatigue', 'tired', 'exhausted' and 'spent'.

There are several sleepiness scales that have been used in fatigue research, such as the Karolinska Sleepiness Scale (KSS), Stanford Sleepiness Scale (SSS) and Epworth Sleepiness Scale (ESS). The KSS is frequently used for evaluating subjective sleepiness (Shahid et al., 2011). The KSS is a nine-point scale ranging from 1 = extremely alert to 9 = extremely sleepy (fighting sleep). A score above 6 is considered sleepy. The SSS is a seven-point scale which measures subjective feelings of sleepiness, e.g. fogginess, responsive but not fully alert, etc. A score above three is considered sleepy. The SSS was developed to quantify the subjective sleepiness of patients throughout the day. In contrast, the ESS measures self-rated average sleep propensity over eight common situations that everyone encounters (Berry et al., 2015).

The KSS and SSS are very similar, both being Likert scales ranging to both extremes of alertness. The KSS has become increasingly popular in drug trial participants, flight crews and other shift-workers (Hirshkowitz and Sharafkhaneh, 2017). Kaida et al. (2006) compared the KSS to electroencephalographic (EEG), electrooculographic (EOG) activity and the PVT to find relatively high correlation between performance measures and subjective sleepiness. Therefore, this research will use the KSS to represent sleepiness as EEG is too invasive for use in field-based research.

2.2.2 Time related fatigue

The circadian rhythm is the body's natural clock, which prompts wakefulness and sleepiness each day. Surgeons work shifts that disrupt circadian rhythms and consequently cause fatigue as a bi-product. The circadian rhythm becomes misaligned in shift-workers when working unnatural hours, with limited or poor sleep beforehand. The circadian rhythm can be adapted to working night-shifts, however the effects may take several days to realign the circadian rhythm to an unnatural cycle. The circadian rhythm is usually suppressed by bright light, which reduces the amount of melatonin the body produces. Melatonin is produced in the evenings as the sun goes down, to dictate when it is time to rest.

However, there are circadian rhythm models, which exist that predict the sleepiness of individuals depending on the time of day. During a 24 hour period, the human body usually experiences two 'sleepy' stages: the middle of the night between 12-6AM and in the afternoon between 2-4PM (see Figure 2.1). Knipling and Wierwille (1994) suggest that the effects of drowsiness accumulate for an hour or more and are associated with a decrease in performance and observable psychophysiological signs.

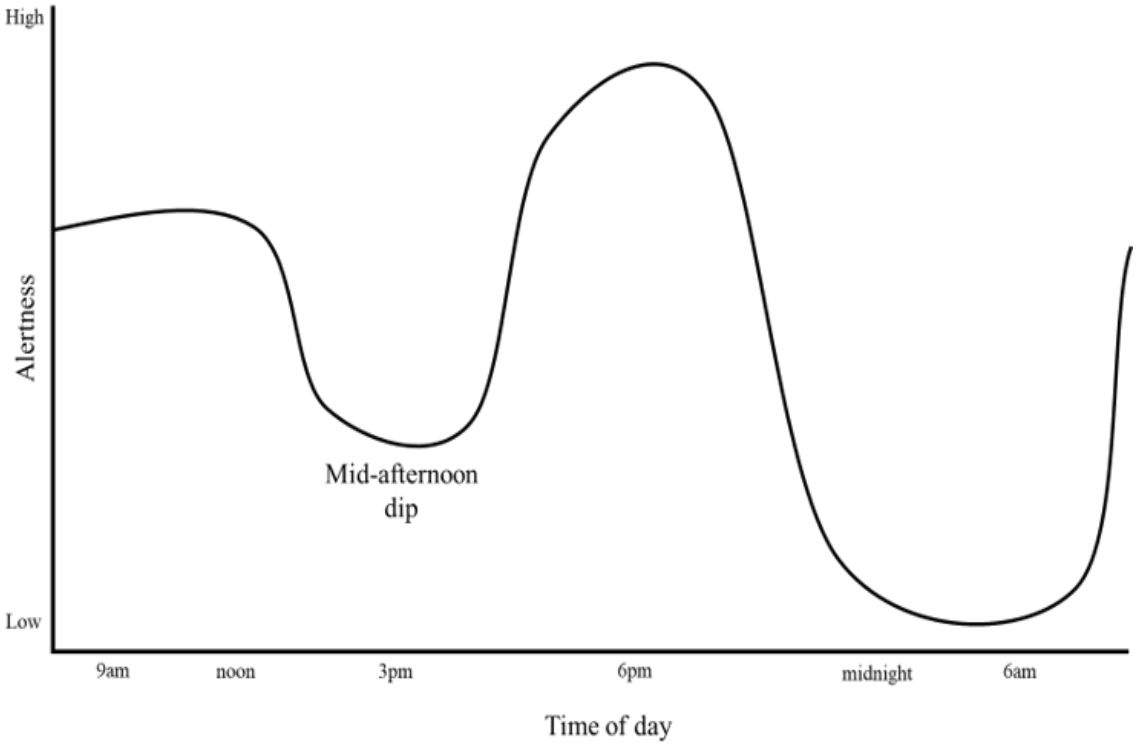


Figure 2.1: Alertness level versus Time of day, showing two dips in alertness during one day, in line with circadian rhythm. (Adopted from Bowman et al., 2012)

2.3 Wearable sensors to monitor fatigue

Wearable technology has become increasingly popular for measuring heart rate variability (HRV) due to its convenience and ease of use. However, when choosing between time-domain and frequency-domain analysis for HRV measurement using wearable technology, there are several important considerations to keep in mind, which are outlined below:

1. Sampling rate: Wearable devices typically sample heart rate at a fixed rate, which is usually in the range of 1-1000 Hz. The choice of sampling rate is important because it determines the accuracy and resolution of the time-domain and frequency-domain analysis. Higher sampling rates provide better accuracy and resolution, especially for frequency-domain analysis, which requires precise measurements of the R-R intervals (Mahdiani et al., 2015). However, higher sampling rates also require more processing power and can drain the battery of the wearable device more quickly.

2. Signal quality: Wearable devices are subject to various sources of noise and artifacts, such as motion artifacts, baseline drift, and electrode contact issues. These can affect the accuracy and reliability of the HRV measurements and can be especially problematic for frequency-domain analysis, which requires clean and accurate signals. It is important to ensure that the wearable device is properly calibrated and that the data is pre-processed to remove any artifacts or noise before analysis.

3. Wearable location: The choice of wearable location can affect the HRV measurement and analysis. For example, wrist-worn devices may introduce motion artifacts that can interfere with the accuracy of the measurements, while chest strap devices may provide more accurate measurements but can be uncomfortable to wear for extended periods of time. The choice of wearable location should be based on the specific research question or clinical application.

4. Processing power: The processing power of the wearable device can also affect the choice of time-domain or frequency-domain analysis. Frequency-domain analysis requires more computational resources than time-domain analysis, and may not be feasible for devices with limited processing power.

2.3.1 Advantages and Disadvantages of different sensors

For the detection of HR and HRV in the lab or field, there are three types of devices currently on the market (Simmonet and Gourvennec, 2016):

- Electrocardiogram (ECG) via electrodes
- ECG via chest belt
- PPG via watch

Each of these devices has its own set of pros and cons, which are discussed below:

Electrodes

Electrodes are typically used in electrocardiogram (ECG) devices, which provide a high level of accuracy in measuring HR and HRV. ECG devices use electrodes placed on the chest, arms, and legs to measure the electrical activity of the heart. However for field based research, this is invasive and expensive, so usually electrodes are placed on only the chest using a 1 or 2 electrodes. The advantages of using electrodes for HR and HRV measurement include:

Advantages:

- High accuracy: ECG devices are considered the gold standard for measuring HR and HRV.

- Wide frequency range: ECG devices can measure HRV across a wide frequency range, from very low to very high frequencies.

- Non-invasive: Electrodes are non-invasive and do not require any invasive procedures.

Disadvantages:

- Inconvenient: The use of electrodes requires the attachment of multiple leads to the body, which can be cumbersome and inconvenient for some individuals.

- Limited mobility: Electrodes may limit mobility during exercise or other physical activities.

- Cost: ECG devices can be more expensive than other wearable devices.

Chest Straps

Chest straps are another type of wearable device that is commonly used to measure HR and HRV. These devices use a sensor that is placed around the chest to measure the electrical activity of the heart. The advantages of using chest straps for HR and HRV measurement include:

Advantages:

- High accuracy: Chest straps are considered to be more accurate than wrist watches for HR and HRV measurement.

- Better fit: Chest straps provide a better fit than wrist watches and can be more comfortable to wear during exercise.

- Affordable: Chest straps are generally more affordable than ECG devices.

Disadvantages:

- Inconvenient: Chest straps can be cumbersome and inconvenient to wear, particularly during extended periods of time.

- Limited mobility: Chest straps may limit mobility during exercise or other physical activities.

- Not suitable for certain body types: Chest straps may not fit properly on individuals with certain body types, particularly those with larger chests.

Wrist Watches

Wrist watches are a popular type of wearable device that can be used to measure HR and HRV. These devices use optical sensors to measure the changes in blood flow that occur with each heartbeat. The advantages of using wrist watches for HR and HRV measurement include:

Advantages:

- Convenience: Wrist watches are convenient and easy to wear, and can be worn all day for continuous monitoring.

- Mobility: Wrist watches do not restrict mobility and can be worn during exercise and other physical activities.

- Affordability: Wrist watches are generally more affordable than ECG devices and chest straps.

Disadvantages:

- Lower accuracy: Wrist watches are generally considered to be less accurate than ECG devices and chest straps for HR and HRV measurement.

- Limited frequency range: Wrist watches are typically limited in their ability to measure HRV at very low and very high frequencies.

- Battery life: Wrist watches may have limited battery life and require regular charging.

For the purposes of fatigue identification in surgeons, PPG via watch is excluded from consideration as surgeons in the NHS are unable to wear any devices below the elbows during surgery. The ECG via electrodes are capable of collecting very long measurements due to the battery life and they do not require Bluetooth for the storage of data, however they require the use of electrodes which can be irritable after prolonged periods of use. Similarly, the ECG via chest belt can also become irritable for extended periods of wear and they are mainly used for extracting information on exercise, nevertheless Simmonet and Gourvennec (2016) identified more reliability and

acceptability for the chest strap when compared to the electrodes, when looking for user acceptability. Umair et al. (2021) found the chest strap to have the largest agreement with a reference device, when looking at time and frequency domain measures of HRV, and HR.

2.3.2 Signal sampling requirements

Heart Rate Variability (HRV) analysis involves measuring the changes in time between successive heartbeats, which are commonly known as R-R intervals. To perform HRV analysis, it is necessary to sample the signal at a sufficient frequency to capture these intervals accurately. The recommended sampling rate for HRV analysis is at least 250 Hz (Mahdiani et al., 2015). This ensures that the R-R intervals can be accurately measured, as well as capturing high-frequency components of the signal that are important for some types of HRV analysis. In addition to the sampling rate, it is important to ensure that the signal is free from noise and artifacts that can interfere with HRV analysis. This can be achieved by using high-quality recording equipment and carefully preparing the subject for the measurement, such as by minimizing movement and ensuring that the electrodes or sensors are properly attached. Overall, to obtain reliable and accurate results in HRV analysis, it is important to adhere to appropriate signal sampling requirements, as well as taking other steps to ensure high data quality. However, in field-based research, artefacts are more prominent due to the increase in movement for participants, as well as putting the equipment on themselves.

Wearable devices have become increasingly popular for monitoring human physiological parameters, such as heart rate, temperature, and physical activity. However, these devices can be subject to various errors and uncertainties that can affect the accuracy and reliability of the measurements they provide. Below, are some of the common errors and uncertainties that arise from wearable devices and their potential impact on data analysis and interpretation.

1. Motion Artifacts

One of the major sources of error in wearable devices is motion artifacts. These are caused by movements of the body, which can interfere with the signals measured by the device. For example, when a person is exercising or walking, the movements of their limbs can cause the device to shift or bounce, leading to inaccuracies in the measurements. Motion artifacts can also be caused by physiological factors such as breathing and muscle contractions. These artifacts can be minimized by using algorithms that remove or correct for them in the data.

2. Signal Drift

Signal drift is another potential source of error in wearable devices. This occurs when the sensor's output changes over time, even when the physiological parameter being measured remains constant. This can be caused by factors such as changes in temperature, humidity, or pressure, or by aging or degradation of the sensor. Signal drift can be corrected by periodically calibrating the sensor or by using algorithms that correct for the drift in the data.

3. Sensor Placement

The placement of the sensor on the body is critical for obtaining accurate measurements from wearable devices. Incorrect sensor placement can result in the sensor not measuring the intended physiological parameter or measuring other physiological parameters that are not relevant to the study. For example, placing a heart rate monitor on the wrist instead of the chest may result in inaccurate heart rate measurements. Sensor placement errors can be minimized by following the manufacturer's instructions and ensuring that the sensor is properly placed on the body.

4. Battery Life

The battery life of wearable devices is also a potential source of uncertainty. When the battery is low, the device may not be able to obtain accurate measurements or may stop functioning altogether. In addition, battery life can affect the sampling rate and duration

of data collection. This can be minimized by using devices with longer battery life or by regularly checking and replacing batteries.

5. Interference

Wearable devices can be subject to interference from external sources, such as electromagnetic fields or other wireless devices. This can lead to errors or inaccuracies in the measurements. Interference can be minimized by using devices that are shielded from external interference or by avoiding areas with high levels of interference. Similarly, for chest straps, contraction of the pectoral muscles provides extensive noise which interferes with the ECG trace, providing unreadable bouts of activity whilst a contraction occurs. This may happen in a variety of circumstances in field-based research, as is the nature of a demanding job.

Wearable devices offer a convenient and non-invasive method for monitoring human physiological parameters. However, it is important to recognize the potential errors and uncertainties associated with these devices and take steps to minimize their impact on data analysis and interpretation. By understanding and addressing these sources of error, researchers and clinicians can obtain more accurate and reliable data from wearable devices, leading to better insights into human health and performance.

2.3.3 Processing in the time domain vs. frequency domain

Heart rate variability (HRV) can be analyzed in both the time domain and the frequency domain. Each approach has its own considerations and advantages, which are outlined below:

<u>Considerations for processing in the time domain:</u>

1. Time-domain analysis focuses on the variability of the R-R intervals, which is the time interval between successive heartbeats. This approach is simple to perform and interpret, making it a popular method for clinical applications.

2. Time-domain analysis provides information on the overall variability of the R-R intervals, which is commonly quantified using standard deviation, variance, and root mean square of successive differences (RMSSD).

3. Time-domain analysis does not provide information on the frequency content of HRV.

Considerations for processing in the frequency domain:

1. Frequency-domain analysis focuses on the spectral decomposition of HRV, which represents the distribution of power in the different frequency bands of the R-R intervals.

2. Frequency-domain analysis provides information on the underlying physiological mechanisms responsible for HRV, such as sympathetic and parasympathetic nervous system activity.

3. Frequency-domain analysis can provide more detailed information on the frequency components of HRV, including low-frequency (LF) and high-frequency (HF) bands. LF is associated with sympathetic activity, while HF is associated with parasympathetic activity.

4. Frequency-domain analysis requires more computational resources and expertise than time-domain analysis.

Overall, the choice between processing in the time domain or frequency domain depends on the specific research question or clinical application. Time-domain analysis is simpler and more straightforward, but frequency-domain analysis provides more detailed information on the underlying physiological mechanisms of HRV. In this research, both time and frequency domain will be analysed.

2.4 Chromatic Monitoring Technique

2.4.1 Introduction

Technological advances are supporting increasing system complexity, requiring a more holistic and intelligent approach for monitoring complex systems (Deakin et al., 2006). The 'Chromatic Monitoring' approach is based upon comparisons, which are translated into mathematical cross-correlations (Jones et al., 2008) involving overlapping procedures, which are matched together (Jones et al, 2000). The approach may involve pure or complex chromatics, i.e. a single signal or multiple sensors with overlapping responsivities. The non-orthogonal nature of chromaticity accommodates a controlled amount of interactivity between receptors (Jones et al., 2008), which can be quantified in terms of amplitude, range and spread of data.

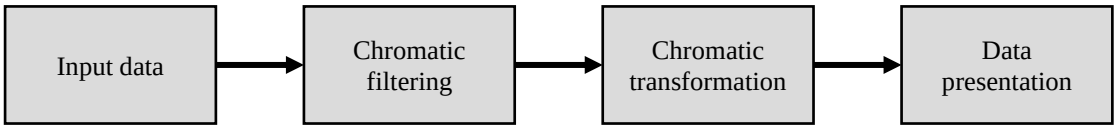


Figure 2.2: Flow chart of Chromatic data processing procedure

2.4.2 Chromatic Monitoring Review

Chromaticity stemmed from polychromatic optical fibre-based sensing in the 1980s but was further developed so that its application could monitor a wider variety of systems (Jones et al., 2000). Processing via non-orthogonal chromatic principles can be used to extract simple, quantitative data from complex signals. Such examples include acoustic monitoring of rail tracks (Deakin et al., 2006), on-line monitoring of contact erosion in high voltage circuit breakers (Wang et al., 2016) and in-vivo monitoring of jaundice in neonate tissues (Sufian et al., 2018).

The chromatic methodology utilises a nonorthogonal system with overlapping receptors, *R, G* and *B* that contain significant levels of information (Jones et al., 2008) from one or more signals, simultaneously. The outputs of these filters are transformed into three chromatic parameters H, L and S which offer signal defining features (Zhang et al., 2004). Once transformed, the data can be represented in polar plots for data recognition.

The chromatic parameters H, L and S are integral components for a holistic understanding in various applications, such as displays, signal processing and image analysis. The HLS (Hue, Lightness and Saturation) provide a versatile representation for achieving precise control over colour properties. H denotes the type of colour, ranging from 0 to 360 degrees, akin to the azimuthal angle. L relates to the perceived brightness and can range on a scale, e.g. 0 (black) to 100 (white). S characterises the intensity or vividness of the colour, varying from 0 (greyscale) to 100 (fully saturated). These parameters offer the benefit of conveying information effectively, providing engineers with a powerful tool to further fine-tune output. For example, in image processing or medical images, adjustment of the HLS can highlight the levels of bilirubin (a yellow compound that forms during the breakdown of red blood cells) in newly born babies, aiding more accurate diagnostics for jaundice (Sufian et al., 2018). This level of control and adaptability afforded by these chromatic parameters significantly advances the capabilities of image processing algorithms.

2.4.3 Chromatic filtering

The overlapping filters *R*, *G* and *B* are generally Gaussian, triangular or truncated in nature (Figures 2.3a, 2.3b and 2.3c). Triangular receptors provide more uniform sensitivity throughout the frequency range of a signal, however Gaussian receptors can provide regions of higher sensitivity, although not uniformly distributed (Wang et al., 2017). The idea of chromatic filtering is that a complex signal structure can be transformed into three values, one for each filter. Generally, each filter has the same width (w) and overlaps uniformly at half the width of the filter. Signal features can be extracted in either time or frequency domains.

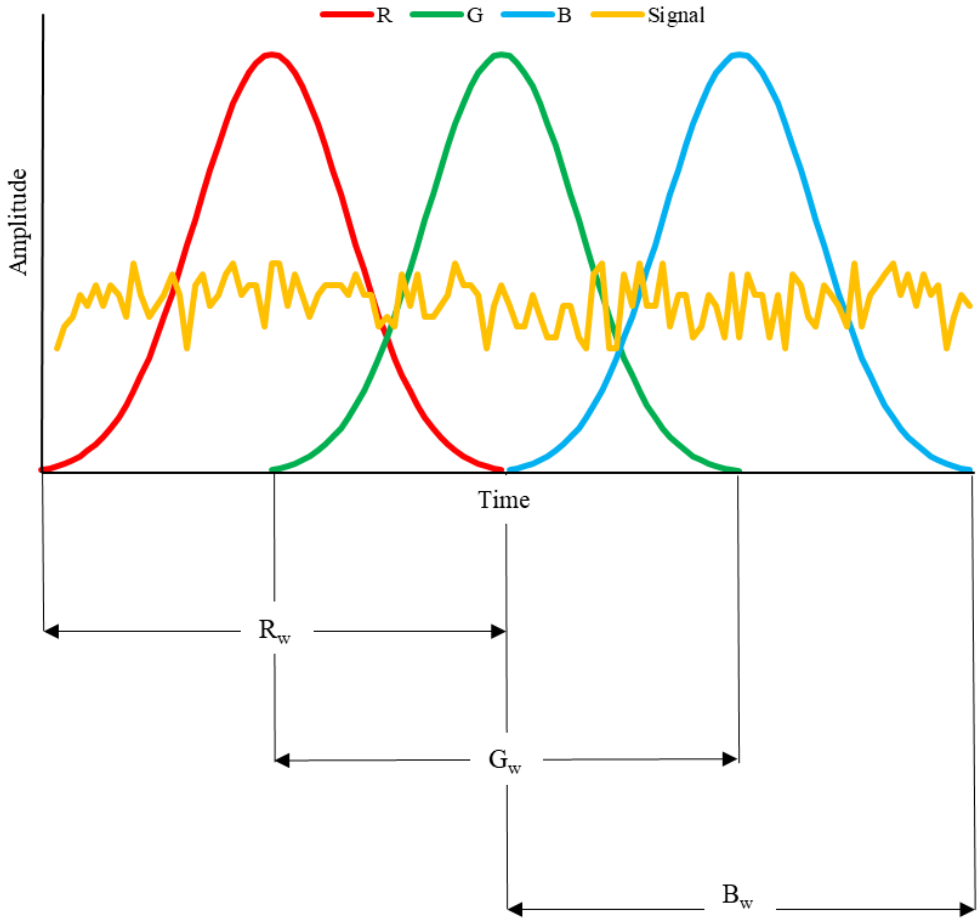


Figure 2.3a: Responses of time domain three nonorthogonal Gaussian filters (R,G,B) superimposed upon a time varying signal.

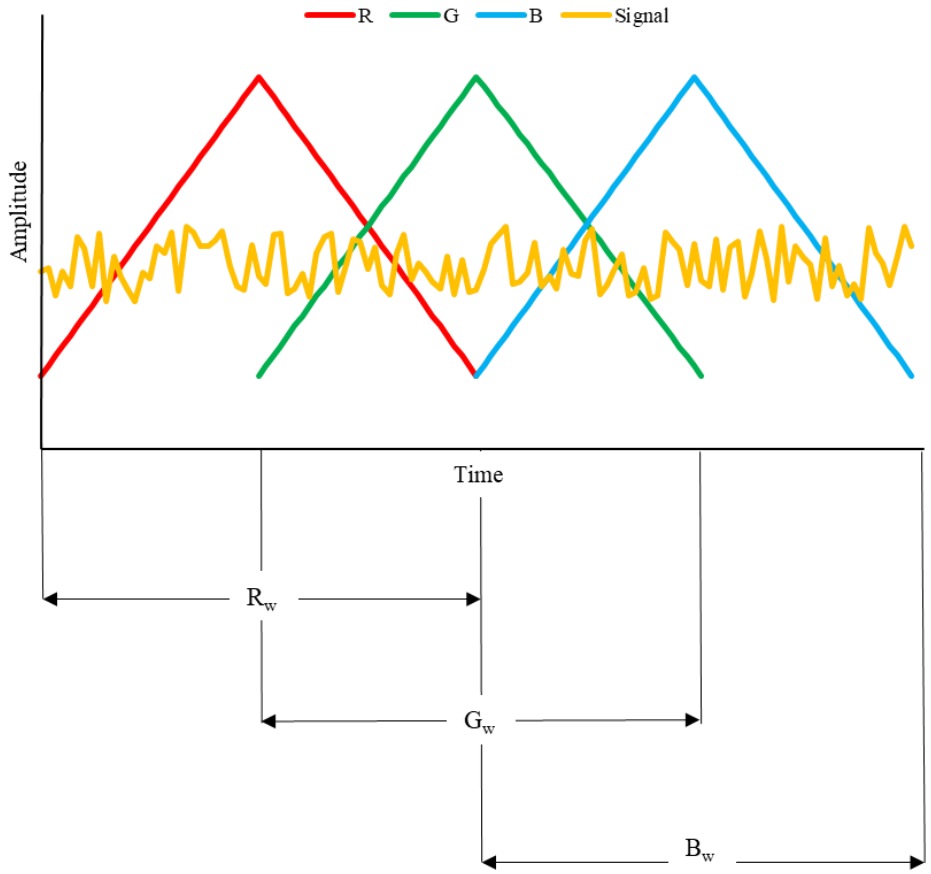


Figure 2.3b: Responses of time domain three nonorthogonal triangular filters (R,G,B) superimposed upon a time varying signal.

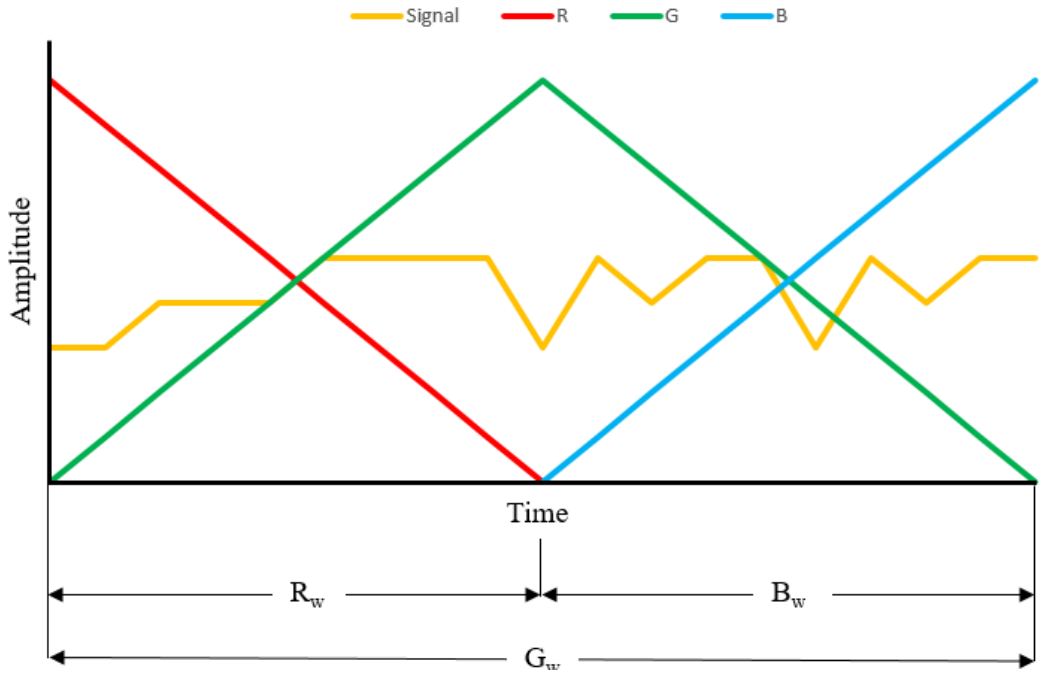


Figure 2.3c: Responses of time domain three truncated filters (R,G,B) superimposed upon a time varying signal.

The vertical axis can for example represent amplitude, voltage or flow rate, depending on the application involved. Additionally, the amplitude and bandwidth of the filters can also be manipulated. However signals must be normalised, particularly if comparing different signals from varying sources simultaneously, onto a common range 0 to 1 (Zhang, 2004). Normalisation creates a common basis for comparison. An equation for normalisation is as follows:-

$$\text{Normalisation of signal} = \frac{\text{Amplitude of signal at one instance}}{\text{Maximal amplitude of signal over entire duration}} \quad \text{----- Eqn 1}$$

Where maximum and minimum values are defined for example HR (220 minus age) and length of shift (e.g. 12 hours). This then enables cross-referencing of signals that will not cause large deviation when calculating chromatic transformation of the signal.

The signal is chromatically filtered after normalisation to get outputs from each filter. Detectors responses R(*P*), G(*P*) and B(*P*) are a function of parameter (*P*). *dP* is the amplitude of a signal that varies with parameter (*P*). *F(P)* is a signal that varies with (*P)*, e.g. time or frequency. The output of a processor $X_o(P)$ (X can be R, B or G) addressing the signal according to Wang et al. (2017):-

$$X_o = \int_P X(P)F(P)dP \quad \text{---------- Eqn 2}$$

2.4.4 Chromatic Transformation

The nonorthogonal outputs (R, G, B) may be transformed into the chromatic parameters *H, L, S* according to the following formulae (Zhang et al., 2004):-

$$H = \begin{cases} 60\left[1+\dfrac{g-b}{max(r,g,b)-min(r,g,b)}\right] & r = max \\ 60\left[3+\dfrac{b-r}{max(r,g,b)-min(r,g,b)}\right] & g = max \\ 60\left[5+\dfrac{r-g}{max(r,g,b)-min(r,g,b)}\right] & b = max \end{cases} \quad \text{---------- Eqn 3}$$

$$L = \frac{(r+g+b)}{3} \quad \text{---------- Eqn 4}$$

$$S = \frac{\max(r,g,b) - \min(r,g,b)}{\max(r,g,b) + \min(r,g,b)} \quad \text{---------- Eqn 5}$$

The chromatic parameter *H* represents the region on the horizontal axis of Figures 2.3a-c, which dominates. Chromatic L represents the strength of the signal, whereby the higher L value the larger the amplitude of the signal. Chromatic S measures the spread of the signal, i.e. a larger spread produces a lower S value (Jones et al., 2000).

2.4.5 Chromatic data representation

The *H, L* and *S* parameters can be represented via coordinates on two-dimensional polar diagrams, *H-L* and *H-S*, where *H* represents the azimuthal angle on the polar diagram and *S* or *L* the radius of the polar plot (Jones et al., 2000). All signals are bound within the polar plot. The importance of this data representation is so signals can be easily distinguished from one another. For instance, a signal obtained by superimposing two separate signals (X_1, Y_1) and (X_2, Y_2) of strengths S_1 and S_2 respectively and separated on a chromatic map by distance l is located at a point given by the simple moment equation (Jones et al., 2000):-

$$S_1 l_1 = S_2(l - l_1) \quad \text{---------- Eqn 6}$$

An example of two different signals which have been chromatically transformed is given in Figure 2.4. The interpretation of the signals represented by *H, L* and *S*

parameters is also given. The two examples of signals are shown, where Signal A has a lower amplitude and narrower spread compared to Signal B. The three overlapping triangular filters are also applied to the diagram. The example shows the parameters where the signal is dominant.

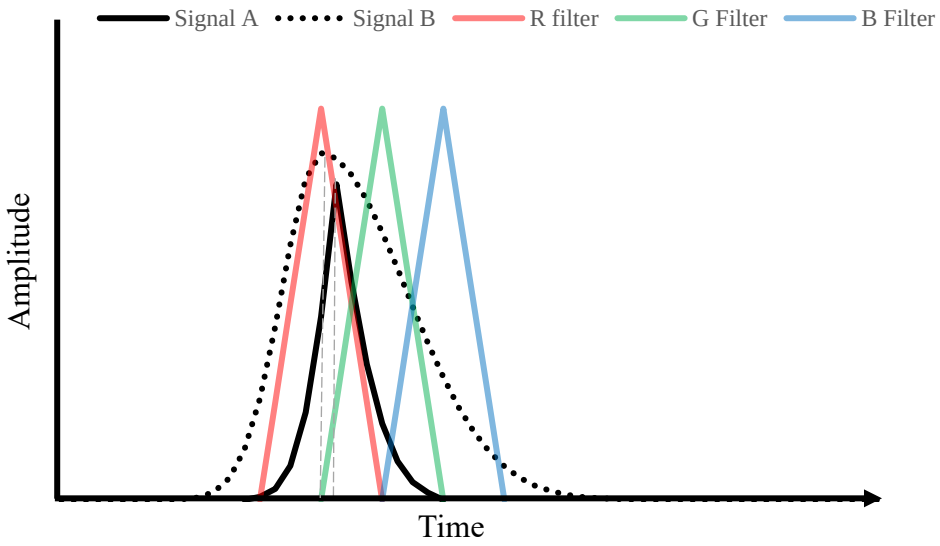


Figure 2.4: Examples of two signals with different spread and strength being chromatically processed over superimposed R G and B triangular filters.

Figure 2.5 shows the H-L and H-S polar diagrams for the two signals presented on Figure 2.4. On these diagrams, signal A is represented by a blue asterix (*) and signal B is represented by a red square (□). Parameter *H* represents in which time region the signal is dominant, which is filter *R* in both cases (Signal A: H = 7.3º, Signal B: H = 40.1º). The chromatic *L* represents the nominal strength of the signal (Signal A: L = 0.15, Signal B: L = 0.44). The chromatic *S* represents the spread of the signal (Signal A: S = 0.95, Signal B: S = 0.39). Therefore, signal B has a higher nominal strength and greater spread with respect to time, compared to signal A.

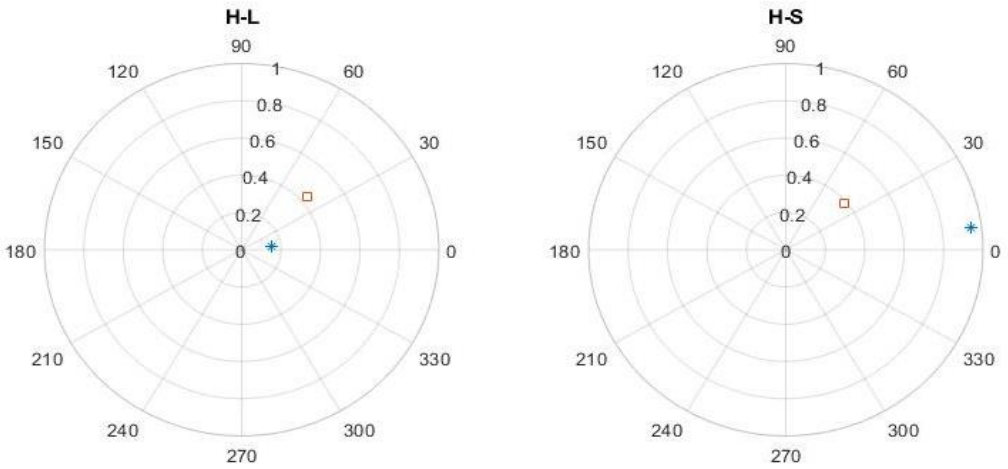


Figure 2.5: Examples of two signals with different strength and spread (*) (□) shown on chromatic maps of (a) H-L and (b) H-S, signals are shown in Figure 2.7.

2.4.6 Time Stepping of chromatics

Each set of filters produces only one set of chromatic parameter values (*H*, *L*, *S*). The filter width (w) is chosen depending on the application. Nonetheless, a process called 'time stepping' is implemented to address the change in a signal over time with the overlapping RBG filters (Figure 2.6). The size of the step may be set in accordance with the application, for example half the width of the filter or if data is averaged over time, every sample of averaged data. Therefore, there is a sequence of chromatic parameter values for each time step. They can then be represented via polar plots (H-L and H-S, Figure 2.8), to show the dominant filter, the signals nominal strength and the spread of the signal over time.

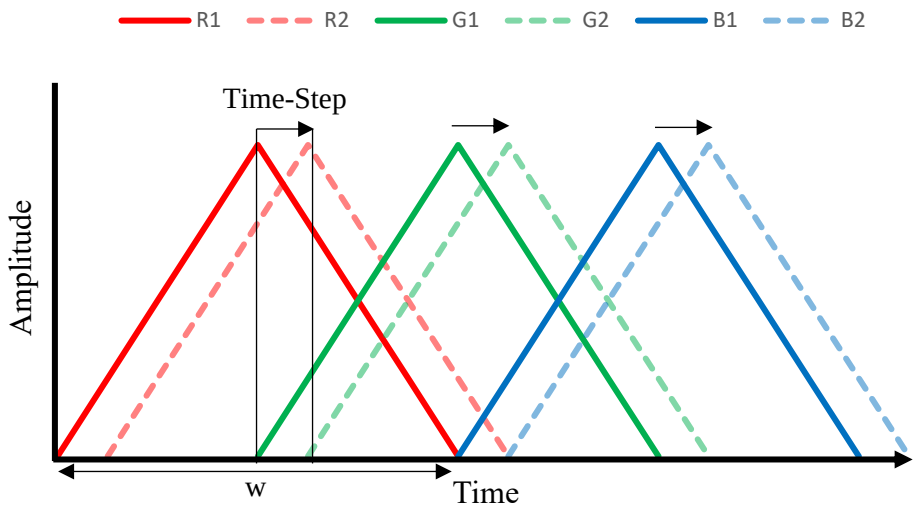


Figure 2.6: Time stepping of R, G, B filters, where the solid triangle filters are the first instance and the dashed filters are the second instance.

An example of time stepping is shown below in Figure 2.7, where triangular filters are displayed in one-minute epochs over a HR signal which is decreasing over time. The corresponding polar plots for the two sets of filters (solid filters and dashed filters) are displayed in the polar plots in Figure 2.8. In this example, the red filter is dominant in both examples.

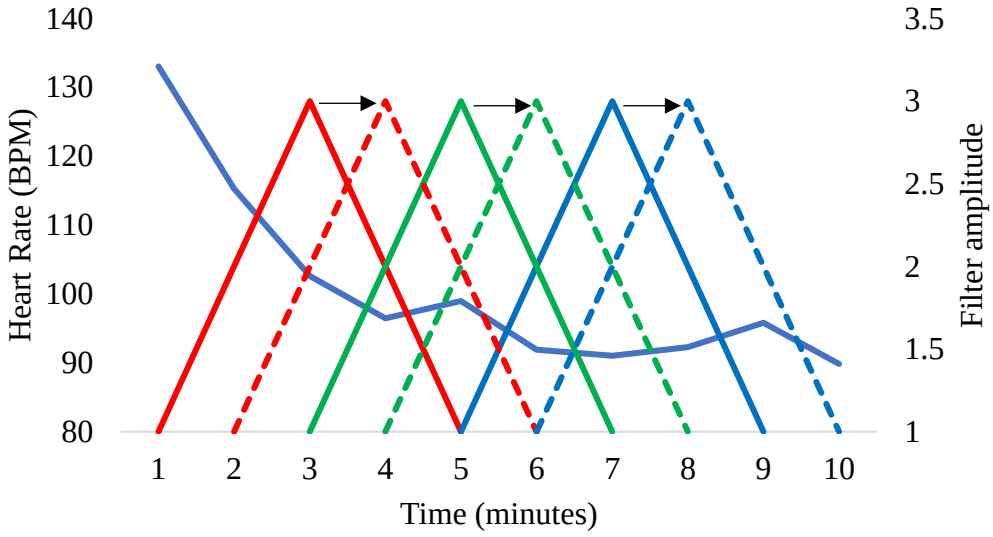


Figure 2.7: Example of HR in series with overlapping time stepping of R, G, B filters, where the solid triangle filters are the first instance and the dashed filters are the second instance.

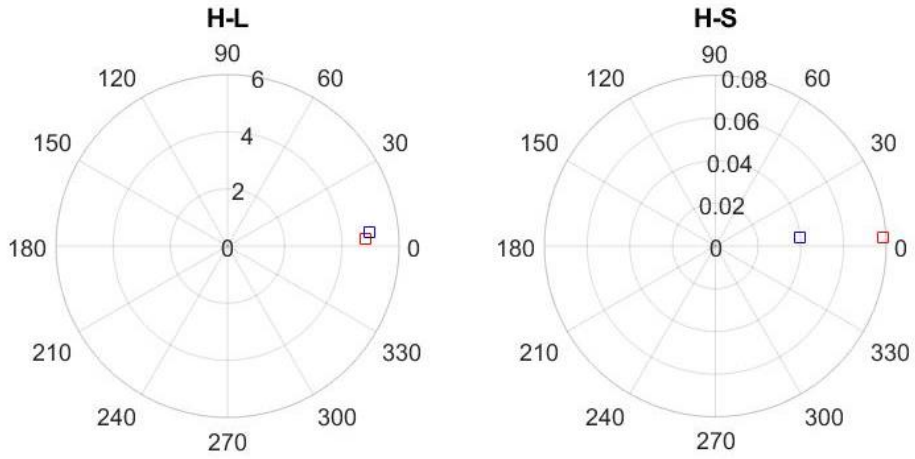


Figure 2.8: H-L and H-S Polar plots of triangular filters in Figure 2.7, where red filter is dominant in both instances. The red squares depict the solid triangular filters and the blue squares depict the dashed filters.

2.5 Summary

This chapter outlines the prevalence of surgeon fatigue and its effects on patient safety. Due to long hours, complex patient care and emotional interactions with patients and families, fatigue in shift workers is significant. Surgical technology is constantly perfected, however the human physiology struggles to keep up, resulting in injuries and/or errors in the workplace. Extensive literature investigates musculoskeletal fatigue in surgeons, however other aspects of a surgical career are seldom reported, such as sleep deprivation or mental exhaustion. Wearable technology must be considered as an alternative solution to obtain objective, clear data sets which may be able to suggest strategical breaks for surgeons.

There are two current methods to monitor fatigue in the workplace, objectively or subjectively. HR can be used for indication of task demand, due to an increased demand for oxygenated blood to the muscles. Similarly, HRV can be used as an assessment of psychological health, as it is associated with regulatory and homeostatic autonomic nervous system recovery. The KSS is frequently used for identifying fatigue in both laboratory and field settings. The body's circadian rhythm prompts wakefulness and sleepiness each day and when misaligned causes fatigue as a bi-product.

The most appropriate method to monitor fatigue incorporating circadian rhythm, hours into shift and HR in the decision making process would be the chromatic approach, offering a holistic approach to identifying the onset of fatigue which can be appropriately individualised. The chromatic methodology can be split into filtering and transformation steps, whereby three overlapping filters sample the signal and then transform to provide the parameters *H*, *L* and *S* (representing dominance, strength and spread of the signal, respectively).

Chapter 3

Preliminary research to find prevalence of fatigue in surgeons

3.1 Introduction

The preliminary, independent research for this thesis is categorised into two parts; a surgeon questionnaire and a pilot study. The surgeon questionnaire (section 3.2) was to identify the contributors to fatigue in the workplace, as well as interventions and mitigation strategies. Of particular interest, was whether there were gender differences associated with surgeon fatigue in the workplace. The pilot study (section 3.3) was a first effort at identifying fatigue in the workplace with the use of wearable sensors throughout the shift. The following section will describe the methodology and results obtained from the preliminary research in both studies.

3.2 Surgeon questionnaire

3.2.1 Introduction

Musculoskeletal fatigue in surgeons is prominent. There is good evidence to support the use of 'microbreaks' (~1 minute pauses every 20-40 minutes with the opportunity to stretch, relax and communicate) in surgery to reduce the likelihood of musculoskeletal injuries and consequently improve patient safety and surgeon longevity (Haramis et al., 2010; Graversen et al., 2011; Engelmann et al., 2011; Dorion and Darveau, 2013; Voss

et al., 2016; Hallbeck et al., 2017; Park et al., 2010). However, fatigue in surgeons goes beyond the operating theatre and limited studies exist monitoring the fatigue of surgeons throughout their shifts, as well as other contributors to fatigue, such as mental exhaustion, compassion fatigue or environmental constructs such as bright lights and/or lack of food during shift (Ulises et al., 2016).

The surgical workforce has changed in recent years, with an increase in female surgeons. Recent studies have evaluated the gender differences in junior residents for perceptions of patient safety, resident education, well-being and job satisfaction (Ban et al., 2021), recognising females as more likely to be dissatisfied with patient safety, overall wellbeing and time allocated with family. Other research aimed at assessing the gender differences in surgeon burnout and psychological wellbeing determined that female surgeons have less rest between shifts, working a higher proportion of >80 hour weeks as well as increased reports of losing sleep over worrying, feeling unhappy and losing confidence in themselves (Dahlke et al., 2018).

Limited evidence exists in the literature for identifying the causes of fatigue at work, as well as any contributors to fatigue, to practically mitigate the onset. Contrary to previous publications, this study aims to identify the gender differences for the causes of fatigue in surgeons within the workplace, understand how fatigue occurs and explore any interventions/strategies that surgeons may apply to try mitigate the effects of fatigue in the workplace. The authors hypothesise heightened levels of fatigue in female surgeons throughout their shift due to the negative work-life balance and heightened levels of compassion fatigue found in previous research (Wu et al., 2017, Dahlke et al., 2018). Drivers for fatigue in female surgeons are hypothesised to be higher perceived levels of demanding work and patient outcomes, and it is hypothesised that younger surgeons will be more susceptible to fatigue than their older, more experienced counterparts (Sturm et al., 2011).

3.2.2 Method

The questionnaire was approved by the University of Liverpool's Ethics Committee on the 30/01/2020. Data was analysed by the research team in line with the University of

Liverpool's ethical guidance. Data was only accessed and analysed by the research team. For data collection, the online questionnaire website 'SmartSurvey' was used to collect anonymous data from surgeons at a hospital in England, UK. Similar to previous research, data was collected from one hospital (Johnston et al., 2018). Participants gave informed consent before participating in the questionnaire. All surgeons working at the hospital were eligible to take part. Surgeon participation was voluntary and they could stop the questionnaire at any point. Surgeons were under no obligation to answer all of the questions.

The 14-part questionnaire (Appendix B) contained four demographic questions; age range, gender, duration of surgical experience and department of work. Estimated time of completion was 5-10 minutes. The questionnaire then asked whether surgeons had been subject to any contributors to fatigue at work (e.g. musculoskeletal, physical or mental exhaustion), what contributes to their fatigue at work and whether they have attempted to minimise these conditions at work (including but not limited to surgery), using a list of multiple-choice answers. For these questions, surgeons were asked to relate to their previous month of work. For the purposes of homogeneity across our sample, other questions included average sleep per night, average hours worked per week, whether they were aware of any intervention strategies to mitigate workplace fatigue and their opinions on taking 'microbreaks' during surgery to reduce fatigue. Additionally, we asked participants their thoughts on wearable technology capable of monitoring their fatigue levels whilst at work and whether they are likely drive a car whilst fatigued. The purpose of these questions was to understand surgeon willingness to have their fatigue levels observed in real time for future research. For all questions, surgeons were given the option to specify 'other', make comments on the question or select the option 'prefer not to say'.

For the purposes of data collection, a liaison at the hospital distributed the questionnaire on behalf of the research team. The survey was first distributed in March 2020 and the survey was closed May 2021. In this time, there were three prompts for surgeons to complete the questionnaire; March 2020, September 2020 and April 2021. Prompts were delivered via email, including a web link to the questionnaire.

Data analysis where appropriate was carried out using SPSS Statistics 24. Relationships between contributors to fatigue and gender were examined by Chi-squared (χ^2) test (Turhan, 2020) (Appendix A). A p-value of 0.05 was used in this research for the interpretation of statistical significance (and throughout the thesis). The p-value is a measure to determine the strength of evidence against a null hypothesis. A small p-value ($\leq$0.05) suggests that the observed data is reasonably likely to occur by chance alone even if the null hypothesis is true. The larger the p-value, the more likely the observed data is to occur by chance alone, even if the null hypothesis is true.

3.2.3 Results

For the questionnaire there was a response rate of 56% (37 out of 66 surgeons completed the questionnaire, with an additional 22 partial completions. Responses included 16 male and 21 female surgeons. Table 3.1 shows the demographic information collected from the questionnaire. No partial completions were included in the final analysis. As shown in Table 3.1, the sample size for some of the categories is small and therefore the results should be interpreted with caution and further research with larger sample sizes is recommended.

On average, 34 surgeons had at least six to eight hours of sleep per night, leaving three surgeons with at least four to six hours of sleep. Tables 3.2 and 3.3 (below) show the contributors to fatigue whilst performing surgery and the interventions to minimise these conditions at work and during surgery (respectively). For hours worked per week, three surgeons worked 30-40 hours, fourteen worked 40-50 hours, fifteen worked 50-60 hours and five worked 60+ hours. Amount of sleep per night and hours worked per week were homogenous for male and female surgeons.

17 surgeons agreed that microbreaks may be a possible intervention to reduce fatigue during surgery. Surgeons were asked whether they were aware of any intervention strategies to mitigate workplace fatigue. 20 were not aware, nine were aware of interventions but do not use them and one did not need to use them as they were not subject to fatigue. The rest used interventions such as; taking rests, yoga, mindfulness activities and disconnecting from work where possible.

Table 3.1: Demographic information of survey takers from surgeon questionnaire.

Age range (Years)	Response total (% respective of gender)	
	Males	Females
25-35	5(31.3)	4(19.0)
36-45	5(31.3)	6(28.6)
46-55	2(12.5)	11(52.4)
56-65	4(25)	0(0.0)
Surgical Experience (Years)		
0-5	2(12.5)	1(3.7)
5-14	3(18.6)	6(28.6)
15-24	6(37.5)	7(33.3)
25-34	3(18.6)	7(33.3)
35-44	2(12.5)	0(0.0)
Surgical Department		
Paediatric	9(56.3)	12(57.1)
Plastics	0(0.0)	3(14.3)
Did not specify	7(43.8)	6(28.6)

Females were statistically more likely be suffering from musculoskeletal injury (20/21 females v 9/16 males, χ^2 (1, N = 37) = 8.15, p = 0.012) and although not statistically significant, from mental exhaustion (18/21 v 9/16, χ^2 (1, N = 37) = 4.00, p = 0.067) in the workplace (Figure 3.1). Females were statistically more likely to allocate their fatigue to prolonged static postures (16/21 v 6/16, χ^2 (1, N = 37) = 5.64, p = 0.018), perceived lack of control on allocated tasks during the shift (13/21 v 4/16, χ^2 (1, N = 37) = 4.98, p = 0.026) and high levels of demanding work (17/21 v 7/16, χ^2 (1, N = 37) = 5.52, p = 0.019) (Table 3.2). To minimise fatigue, females were statistically more likely to change positions during their shifts (15/21 v 5/16, χ^2 (1, N = 37) = 5.90, p = 0.015) (Table 3.3). Females were more likely to have discussion with supervisors over their level of workplace fatigue (6/21 v 0/16, χ^2 (1, N = 37) = 4.41, p = 0.057), though this is not statistically significant. All other fatigue contributors, causes of fatigue and interventions were homogenous across our sample. The analysis of this data is derived from one set of data.

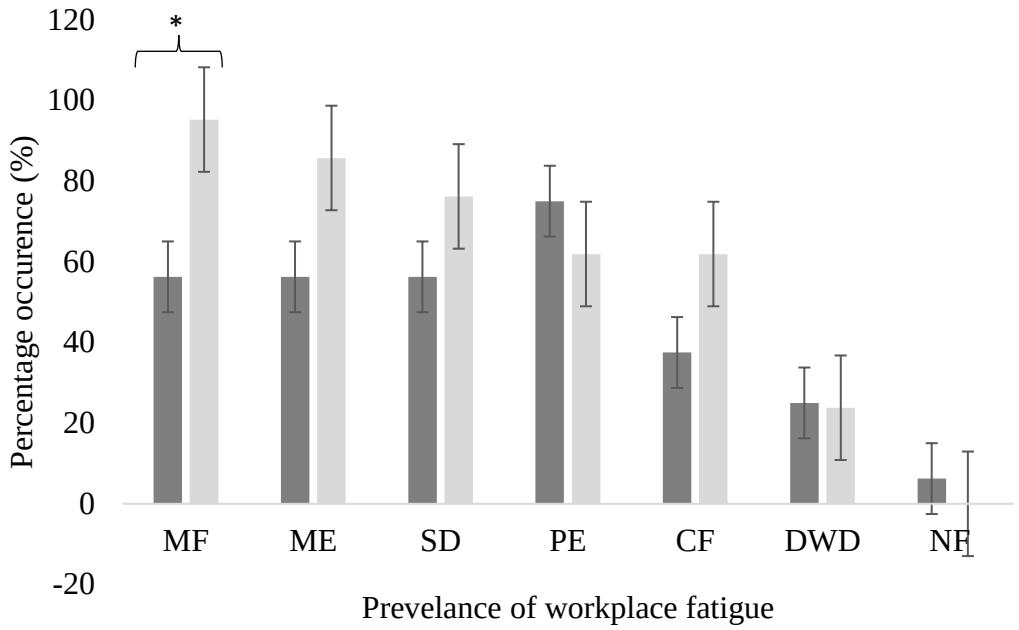


Figure 3.1: Prevalence of workplace fatigue split by gender in descending order for females with error bars (MF: Musculoskeletal fatigue, ME: Mental exhaustion, SD: Sleep Deprivation, PE: Physical Exhaustion, CF: Compassion Fatigue, DWD: Drowsiness whilst driving, NF: No Fatigue). *p = <0.05.

We asked the surgeons their thoughts on wearing a base-layer vest capable of monitoring their fatigue levels during their shift. We found fifteen surgeons were inclined to wear the vest, whilst seventeen are either hesitant or would need more information. The remaining five surgeons would not wear the vest. Finally we asked if surgeons found themselves fatigued after a shift, whether it would influence their decision to drive home, to which we found, though five participants said they were aware they should not drive, twenty-one would drive straight home, eight would take a break/nap before driving home and the rest do not drive.

Table 3.2: Contributors of existing fatigue at work/whilst performing surgery categorised by gender in descending order for females.

Causes of fatigue at work/whilst performing surgery	Response total (% respective of gender)	
	Males	Females
High levels of demand	7(43.8)	17(81.0)*
Persistent mental activity	8(50)	16(76.2)
Prolonged static posture	6(37.5)	16(76.2)*
Lack of control at work	4(25)	13(61.9)*
Lack of sleep	7(43.8)	12(57.1)
Lack of rest during shift	6(37.5)	12(57.1)
Uncomfortable posture	6(37.5)	12(57.1)
Table height	5(31.3)	12(57.1)
Lack of food/water	5(31.3)	12(57.1)
High levels of effort	5(31.3)	11(52.4)
Length of operation	6(37.5)	10(47.6)
Patient outcomes	5(31.3)	10(47.6)
Disputes with colleagues	2(12.5)	9(42.9)
Bright lights	2(12.5)	8(38.1)
Positioning of apparatus	5(31.3)	7(33.3)
Physical exertion	4(25)	7(33.3)
Lack of reward/recognition	3(18.6)	7(33.3)
Monitor height	2(12.5)	7(33.3)
Continual tension using hands	2(12.5)	7(33.3)
Discussion with friends/family of patient	4(25)	4(19.0)
Lack of motivation	2(12.5)	4(19.0)
No fatigue	1(6.3)	0(0.0)

*p = <0.05

Table 3.3: Interventions to minimise fatigue at work/during surgery categorised by gender, in descending order for females.

Interventions to reduce fatigue	Response total (% respective of gender)	
	Males	Females
Take a break when possible	11(68.8)	17(81.0)
Change positions	5(31.3)	15(71.4)*
Adjust the surgical field	5(31.3)	13(61.9)
Relaxation/exercise techniques (stretches, etc.)	5(31.3)	12(57.1)
Change height of surgical field	5(31.3)	8(38.1)
Treatment/medication	1(6.3)	7(33.3)
'it's impossible', 'it's part of the job'	5(31.3)	6(28.6)
Change instruments	1(6.3)	6(28.6)
Ignore it	3(18.6)	5(23.8)
Discussion with supervisors	0(0.0)	5(23.8)
Time away from operating room	1(6.3)	3(14.3)
No fatigue	1(6.3)	0(0.0)

*p = <0.05.

Table 3.4: Comments for microbreak and wearable vest acceptance, split by gender.

	Comments
Microbreaks	
Female	*Enough time lost from list by staggered admissions to day case lists so less operating time.*
Male	*I feel that it is the other demands of the job that lead to fatigue rather than just operating.*
	Noted in long operations when I did take a break then more readily able to complete the procedure.
Wearable vest	
Female	*I'd like to know how reliable it is at giving the right information.*
	Conflicted about this. Feel wary about other team members commenting on my (personal) fatigue levels, but at the same time have witnessed surgeons in the past refusing to take a break and struggling with surgery, leading to complications.
	Don't like the thought of a vest. Would be happy to wear a watch/monitor.
	BUT always too hot.
	I get really hot.
	I would prefer to wear an ankle monitor or something that would not make me hotter/sweatier during operating!
Male	*I am not sure that being more aware of fatigue would help if there is no option to hand over to a colleague.*

3.2.4 Discussion

Females more commonly reported the presence of musculoskeletal fatigue than males. This seemingly develops from prolonged static postures, which requires changing positions at a greater frequency than males. This finding may identify that the operating theatre is more suitably designed for the male stature. Females may benefit greatly from the use of microbreaks when compared to males and future research should look at the gender-differences of intra-operative fatigue.

Similarly, females were more likely to suffer from mental exhaustion than males, which could be attributed to higher levels of perceived demand and lower levels of control over their workload in the workplace (Table 3.2). This finding may coincide with previous literature and the negative work-life balance which females more frequently experience (Sturm et al., 2011, Wu et al., 2017). Since these findings only consider half of the surgical cohort answered the questionnaire, this result should be analysed cautiously. However, whilst only a proportion of females would discuss their levels of

fatigue with supervisors, there is evidence that significant stigma is attached to fatigue and thus reluctance to discuss fatigue with supervisors. It is previously addressed that males perceive to have more consistent leadership support than females, expressing concern that there may be gender bias encountered in the workplace, or inadequate support for females and their negative work-life balance (Wu et al., 2017). Nevertheless ninety-seven percent of our cohort suffer from fatigue however only sixteen percent discuss fatigue with their supervisors. Evidence to explain why surgeons do not adopt intervention strategies is limited. One previously identified reason may be the 'feeling of invincibility' which can be accounted for by thirty percent saying 'it's part of the job' to encounter fatigue (Table 3.3) (Dorion and Darveau, 2013). Whilst it is understandable that rigorous training prepares surgeons for complex shifts, complex patient interactions and quick decision making, workplace fatigue can easily cause sleepiness and impaired judgement, which if ignored, leads to compromised patient safety.

Whilst female surgeons had a greater amount of contributors and interventions to mitigate fatigue in the workplace, males currently suffering with musculoskeletal injury were significantly more likely to adopt microbreak practices and wear a vest capable of monitoring fatigue levels at work, showing promise for the market of wearable technology in the future. Surgeons had the option to write comments for microbreaks and base-layer technology (Table 3.4). When wearable vest acceptance was split by gender, it is clear to see there are more concerns for females wearing the vest than males, regarding temperature in the workplace. For microbreaks, one male participant demonstrated knowledge of the benefits of adopting them, however one female participant suggested there is not enough time to adopt them due to staggered admissions. As such, more research is required into the gender differences of user acceptability for the possible widespread application of microbreaks and wearable technology into the workplace.

In order to mitigate the onset of fatigue, acknowledgement of these gender differences in fatigue are critical to reduce the statistical likelihood of fatigue onset in both genders (Lu et al., 2020). Making fatigue challenges overt through support programmes may help strategize shift patterns and improve work-life balance to reduce the onset of preventable errors made in the workplace to improve patient safety.

There are limitations to this research analysis. Firstly, we can only base our research on the sample of data we have collected. The questionnaire has twenty-two partial completions. Surgeons who are currently suffering with fatigue may have an increased likelihood to respond to the questionnaire resulting in bias.

3.2.5 Conclusion

The findings indicate a prevalence of gender-based differences of fatigue in the workplace for surgeons, as well as the key contributors to fatigue. The study identified an increase in the prevalence of mental exhaustion and musculoskeletal fatigue in females, possibly due to perceived lack of control over workload, high levels of demanding work and prolonged static postures at work. Our findings indicate that workplace fatigue is prevalent across all demographics and mitigating fatigue from the start of surgical careers could help reduce the onset of drivers of fatigue, such as musculoskeletal injuries, mental exhaustion and time missed from work as a result of workplace fatigue. Future studies may consider why females are reporting higher levels of demanding work and lower levels of perceived control over work, as this may directly impact the level of mental exhaustion felt by so many female surgeons.

3.3 Pilot study

3.3.1 Introduction

As aforementioned, biomarkers such as HR, HRV and skin/core temperature (ST, CT) have been previously used to successfully identify fatigue in the workplace (Johnston et al., 2018, Pichot et al., 2002). Wearable technology can provide thorough insight into one's physiology and can identify the onset of fatigue in individuals. To date, no studies have monitored surgeon fatigue throughout their shifts to see what predisposes surgeons to fatigue and whether these can be mitigated. Monitoring fatigue in real time may be able to recommend to surgeons the opportune moment to take a break during their shift to reduce the likelihood of errors. The possibility of identifying fatigue through the use of wearable technology offers a novel approach to the problematic effects of workplace fatigue.

This present study combines a previous methodology (Johnston et al., 2018) along with updated techniques using surgeons as our demographic, to see what predisposes surgeons to fatigue whilst at work. Similarly, no studies have observed the differences in objective and subjective fatigue levels across different shift types; day, night or 24 hour shifts in surgeons. Additionally, the present study will add extra biomarkers including an objective marker of sleepiness (psychomotor vigilance test) and a skin/core body temperature sensor.

The aim of this study is to cross-reference questionnaires and biomarkers throughout a surgeons' shift to;

i) Use a linear mixed models approach to identify any significant effects of variables on self-perceived sleepiness.
ii) Examine the differences in fatigue across different shift types.
iii) Examine age differences related to fatigue.

3.3.2 Method

Participants were recruited through a liaison at the hospital (UK) for data protection purposes. Once a suitable shift was chosen, the primary investigator placed the equipment in a safe location at the hospital for the surgeon to collect 15 minutes before their shift was due to start (in line with COVID guidelines at the time of data collection). Surgeons wore the CORE body temperature sensor and Actiheart monitor before completing a psychomotor vigilance test at the beginning of their shift. Following on, every 4 hours (including the start and end of the shift), surgeons were prompted by automated text message to fill out a questionnaire. Simultaneously, a link to the Surgical-Task Load Index (SURG-TLX) was included in the text message to be completed after an operation (if appropriate). At the end of the shift, the equipment could be removed and a final psychomotor vigilance test was completed. Data was collected between February and May 2021.

21 surgeons (15 male, 6 female) were recruited for the study. Participants gave informed consent before participating in the trial. The study was approved by the University of Liverpool's Ethics Committee. Data was analysed by the research team in line with the University of Liverpool's ethical guidance and in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments. Data was only accessed and analysed by the research team. Overall, 9 day shift (8am-8pm), 9 night shift (5pm-8am) and 4 twenty-four hour shift (8am-8am) questionnaires were analysed. Four participants were aged 35-44, the rest were aged 25-34. Mean surgical experience (years) was 6.41 (±4.06). Overall, 15 reaction time tests were analysed (11 male, 4 female) and 9 complete sets of objective data (7 male, 2 female) were collected. Missing data was due to technical errors or limited time before the shift to set up the equipment.

Participants wore an ECG via electrode device (Actiheart, 2021) to measure activity (measured in counts) and heart rate. Data was extracted in 1Hz samples throughout the shift. The Actiheart is fitted with two electrodes, one on the sternum and one on the left side of the chest. Heart rate variability was analysed using standard deviation of N-N intervals.

Participants wore a temperature sensor (GreenTEG, 2021) to measure skin temperature and core body temperature simultaneously. Similar to the ECG electrode device, 1Hz data samples were collected throughout the shift. Participants were required to download the CORE application in order to start recording the data. The primary investigator downloaded the data once the equipment was collected for analysis. The sensor was worn using a chest-strap and the sensor placed underneath the left arm.

A psychomotor vigilance test was custom made using an Arduino Uno fitted with an SD card and a battery for portability. The reaction time test had two buttons; one to start the test and one to react to a visual stimulus. The psychomotor vigilance test lasted for three-minutes, where participants reacted to a light turning on as fast as they could

by pressing a button. The light would turn on at random intervals between 2-5 seconds. Participants were asked to complete the reaction time just before the start of their shift and at the end of their shift, as an objective measure of their current fatigue level. Minor lapses in concentration were between 350-500ms and major lapses were identified as >500ms (Basner, 2011).

Every 4 hours (including the start and end of the shift) or when most convenient, participants were asked to complete a questionnaire. The questionnaire was administered via automated text messages containing a hyperlink to a webpage. The questionnaire was created using JISC online surveys. The questionnaire included Likert scales from 1 (not very) to 10 (extremely). Questions included levels of; sleepiness (using the KSS (Shahid et al., 2011)), control, motivation, demand, mental effort and reward on shift, with the addition of capability to use fine motor skills, capability to make the right decision and whether they had made an error or near-miss error since the last questionnaire.

The SURG-TLX (Wilson et al., 2011) was created to develop and validate surgery-specific workload measures. Participants were instructed to fill out this questionnaire after an operation, at the point they remember feeling the most challenged. Again, Likert scales from 1 to 10 were used to decipher; how mentally/physically fatiguing, the pace, the complexity, how anxious the surgeon was, how distracting the operating theatre was and how sore/stiff the surgeon felt (including anatomical location if appropriate). Additionally, duration and time of procedure was recorded.

For some participants, artefacts occurred with the ECG via electrode heart rate monitor. In such instances, MATLAB's live script feature, 'clean outlier data' was used to linearly interpolate missing values and remove values that were recorded as 0's for heart rate. To reduce the amount of samples to perform statistical analysis on, data was averaged into 5-minute samples (i.e. the average of every 300 samples). All statistical analyses for biomarkers were analysed using 5-minute averages. Resting heart rate of

the individuals was identified using at least 10 minutes of data during the shift where activity was '0' whilst they were not operating, to account for within-subject heart rate variability and calculate heart rate percentage above rest for Linear Mixed Model (LMM) analysis.

SPSS was used for data analysis of the study. For analysis of the reaction time, paired sample T-Test identified mean differences pre and post shift. T-tests and Analysis of Variance (ANOVA) tests (Thomas et al., 2011) were used to compare means between shift-types and look at the change in variables pre-post shift. Statistical analyses are explained in more detail in Appendix A. Where data breached parametric assumptions, non-parametric tests were conducted.

LMM (Bryk and Raudenbush, 2002) identified whether any variables effect the KSS variance. LMM analysis only included participants with full data sets (nine participants in total). Incorporated into the LMM model; a random intercept for subject to account for within-subject correlations and core temperature, skin temperature, movement, hours into shift, time of day (24hrs), standard deviation of N-N intervals and heart rate percentage above rest as fixed effects. The LMM accounts for correlation between repeated measures and within subjects and allows for a separate time trend for each subject. Therefore, the LMM can be used to calculate how biomarkers effect KSS over time.

3.3.3 Results

Paired sample t-test identified no significant differences between pre-post shift for reaction time, minor lapses or major lapses ($t(14) = -1.089$, $p = 0.295$, $t(14) = -1.586$, $p = 0.135$, $t(14) = -.702$, $= 0.494$, respectively) despite a significant difference in self-perceived sleepiness pre-post shift ($t(14) = -5.522$, $p < 0.001$). There were no significant differences between genders and shift types. The average KSS scores across shift types are depicted in Figure 2.2 (section 2.2.2).

For the SURG-TLX, one-way ANOVA identified no mean differences between shift types, age groups or experience. Pearson's correlation (Freedman et al., 2007) identified no correlation between KSS and years of experience. However there was a strong correlation between the duration of surgery and mental fatigue, physical fatigue and

complexity of surgery ($r(16) = 0.65$, $p = 0.006$, $r(16) = 0.74$, $p = <0.001$, $r(16) = 0.79$, $p = <0.001$, respectively). A strong correlation was identified between anxiety and distractions in the operating theatre ($r(16) = 0.76$, $p = 0.001$). Sixty-three percent of surgeons experienced some level of musculoskeletal soreness to either the neck, arms or back. There was no significant difference between duration of surgery and musculoskeletal soreness ($r(16) = 0.49$, $p = 0.054$).

The surgeon questionnaire found no strong correlation between sleepiness and control, demand, motivation, mental effort or reward. However, a moderate, negative correlation was identified for sleepiness and motor skill capacity ($r(73) = -0.43$, $p < 0.001$) and decision making ($r(73) = -0.38$, $p = 0.001$). A strong correlation was identified between mental effort and demand ($r(74) = 0.78$, $p < 0.001$) and reward and motivation ($r(74) = 0.62$, $p < 0.001$). Our questionnaire found no reports of errors made from all surgeons, with three instances of near-miss errors on night shifts for surgeons aged 24-34. However, Mann-Whitney U analysis (McKnight and Najab, 2010) identified greater self-perceived control, greater motor skills capability, less demanding shifts and less mental effort on shifts for 35-44 year olds compared to 24-34 year olds ($U = 236$, $p = 0.020$, $U = 195$, $p = 0.004$, $U = 120$, $p < 0.001$, $U = 237$, $p = 0.020$, respectively).

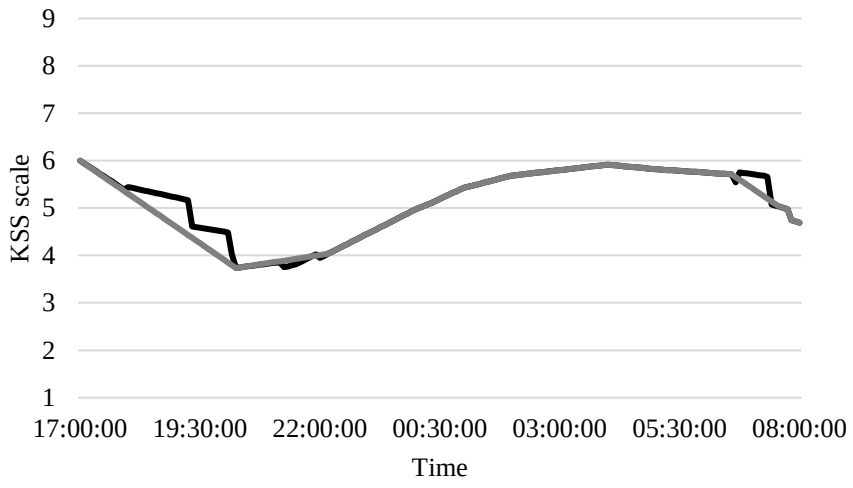
One-way ANOVA (Moore and McCabe, 2003) identified significant differences between shift type and self-perceived control, motivation, demand, mental effort and reward. Overall, 24 hour shift workers report more control over their shift than day and night shift workers ($F(2,73) = 4.19$, $p = 0.019$). Night shift workers reported less motivating and less demanding shifts than 24 hour and day shift workers ($F(2,73) = 5.73$, $p = 0.005$, $F(2,73) = 13.29$, $p < 0.001$, respectively). Additionally, night shifts were less rewarding and required less mental effort than day shifts ($F(2,73) = 5.25$, $p = 0.007$, $F(2,73) = 3.61$, $p = 0.032$). Overall, day-shift workers were significantly less tired than night-shift and 24-hour shift workers ($F(2,1085) = 65.65$, $p < 0.001$).

LMM analysis identified a reduction in heart rate percentage above rest ($F(1,1076) = 7.84$, $p = 0.005$), core temperature ($F(1,1073) = 13.85$, $p <0.005$) and skin temperature ($F(1,1079) = 9.94$, $p = 0.002$) to significantly increase self-perceived sleepiness (KSS). Time of day ($F(1,6) = 3.97$, $p = 0.094$), standard deviation of N-N intervals ($F(1,1075)$

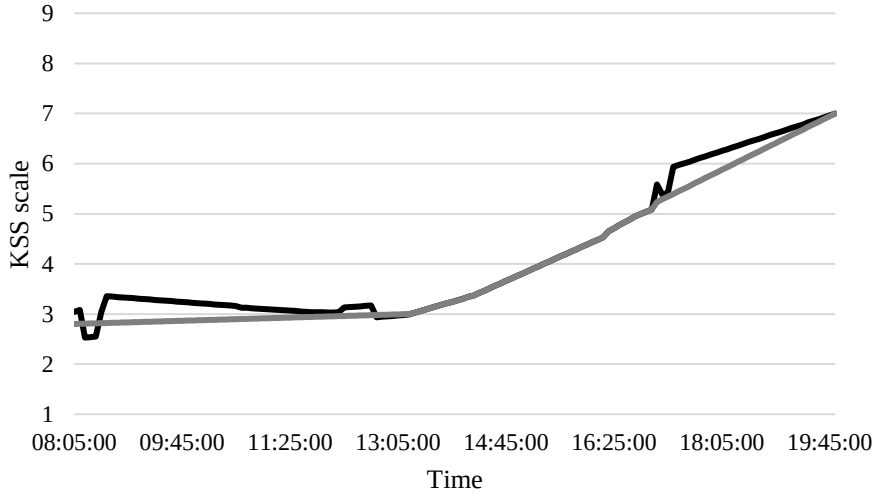
= 3.52, p = 0.061), hours into shift (F(1,6) = 3.33, p = 0.120), movement and gender (F(1,1073) = 1.48, p = 0.224, F(1,6) = 0.66, p = 0.449, respectively) were not significant.

This pilot study identified KSS trends: for day, (8am-8pm, n=9), night (5/8pm-8am, n=9) and 24 hour (8am-8am, n=4) shifts, where n = number of participants (see Figure 3.2 below). The night shift can be split into two different shifts, either 5pm or 8pm starts.

a)



b)



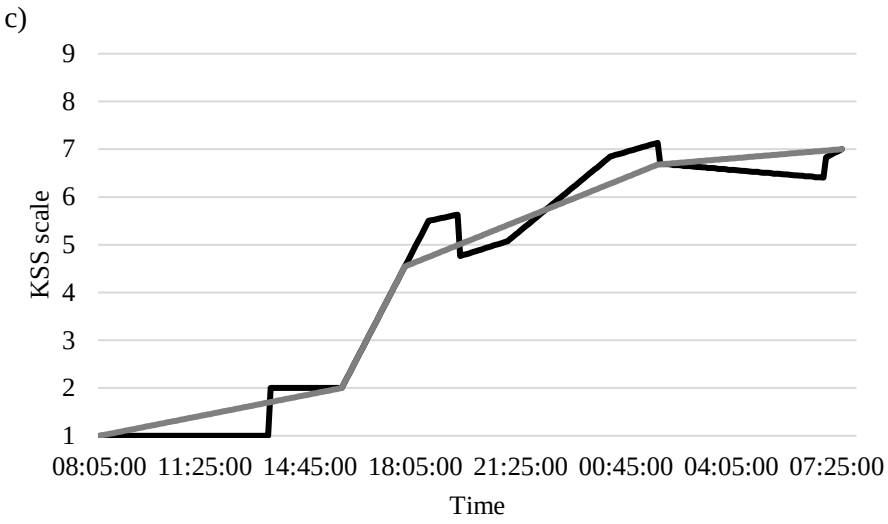


Figure 3.2: Karolinska sleepiness scale (KSS) versus Time of day (a) Night shift, (b) Day shift, (c) 24hour shift (black line is original, grey line is smoothed).

The figures show the average sleepiness on shift for the three different types of shifts the surgeons may take. The black line is the original data. The grey line is the data smoothed for the purposes of the algorithm in this study used to predict fatigue and also to remove outliers, mitigating sudden changes in the dominant chromatic filter (section 2.4).

The night shift graph shows fatigue for those who began their shifts at 5pm, however by 8pm the sleepiness levels off. This could be due to a lack of sleep in the participants who began at 5pm. However, sleepiness increases until it reaches its apex at around 4am, in agreement with the 24hr circadian rhythm model. At this point, the surgeons will be starting to feel tired. The curve then tapers off at around 6am until the end of the shift, which could be due to the sunrise, suppressing the production of melatonin.

It can be seen in the day shift graph, that time on shift significantly effects sleepiness (Johnston et al., 2018). At around 1pm, sleepiness starts to rise until the end of the shift, similar to the mid-afternoon dip. However there is no decrease in sleepiness observed until the surgeons clock-off.

For the 24 hour graph, surgeons are particularly well rested at the beginning of the shift, with an increase in sleepiness from around 1pm until 6pm. Sleepiness continues to rise however, due to the period of time staying awake on shift. Surgeons working 24 hours with minimal breaks are therefore at a heightened risk of fatigue.

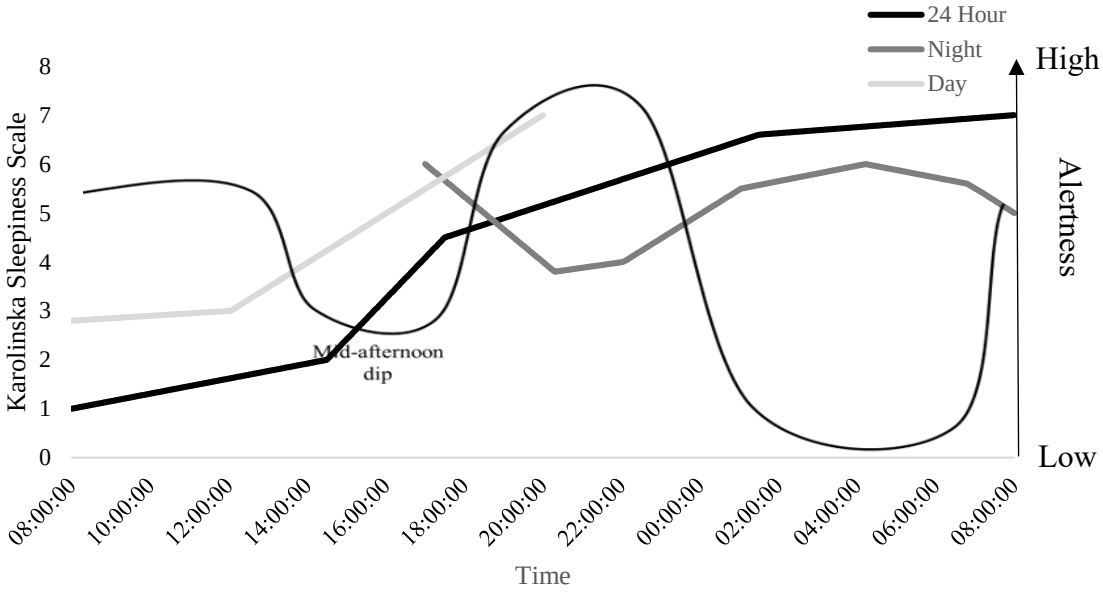


Figure 3.3: Figures 2.1 and 3.2 superimposed to show the inverse relationship between the established circadian rhythm and KSS scores. (Adopted from Bowman et al., 2017).

Figure 3.3 shows the roughly inverse relationship between the established circadian rhythm and the KSS scores from 21 surgeons, providing validity of the predicted KSS model in the algorithm. Jasper et al. (2009) also found a similar circadian rhythm trend for >24-hour sleep deprivation every three hours when comparing KSS to salivary melatonin production (used to monitor the circadian rhythm).

3.3.4 Discussion

Growing literature is indicative for fatigue effecting medical errors in surgeons (McCormick et al., 2012). Our analysis is suggestive that wearable technology may serve as a proxy for the onset of fatigue identification and thus the mitigation of fatigue in the workplace to conceivably reduce medical errors. For the onset of physical fatigue, increases in heart rate and temperature generally increase the onset of fatigue (Aryal et

al., 2017). However, our analysis identifies the opposite trend for the onset of fatigue in surgeons, in a workplace setting. This difference is associated with the definition of fatigue, whereby physical fatigue is associated with diminished capacity to perform tasks due to exertion, whereas workplace fatigue is associated with a combination of physical exertion and mental stress. HR is generally higher for physical jobs which causes fatigue quicker due to energy expenditure and continuous physical activity fatigue (Aryal et al., 2017). In the surgical workplace, whilst high HR is identified, it is less frequently observed and this study identified fatigue was more prominent at quieter periods during the shift, perhaps during monotonous work. The present study identifies significant effects of heart rate percentage above rest, skin temperature and core temperature on self-perceived sleepiness from LMM analysis, however a small sample size of nine surgeons were included in the analysis. A reduction in heart rate, skin and core temperature increases Karolinska sleepiness Scale scores.

Despite this, the individual weightings of these significant variables are minimal and account for only a small variation in the KSS. Variables such as time of day and hours into shift provided greater variability for the fluctuation of KSS, but were not significant in doing so. Table 3.5 below shows the individual weightings of each variable (by each unit of '1'), and the effect they have in altering the KSS. The numbers in Table 3.5 were derived from the LMM analysis. For example, working for 5 hours would give the calculation: (5x1) x 0.183, which would increase the baseline KSS by 0.915. If an individual started their shift by answering a '3' for the KSS, then after 5 hours their KSS would be 3.915. If an individuals resting HR was 60bpm and their HR raised to 75bpm, then: (15x1) x -0.007, which decreases the KSS score by 0.105. With all the variables included, a KSS score can be given which is indicative of their temperature, HR, time of day (e.g. day or night shift), etc.

Undoubtedly, surgeons do not have enough time between shifts to acclimatise fully to night shifts and therefore this natural reduction in core body temperature correlates with an increase in the Karolinska sleepiness Scale (Arendt, 2010). The reduction in heart rate and skin temperature may be more apparent on night shifts due to the reported reduction in demand and mental effort when compared to day shifts. These instances at

work may occur with low-impact, monotonous tasks such as filing reports or analysing medical records. Suggestively, times during the shift where there is little engagement, motivation and interaction with external stimulus, are the most likely times for surgeons to report high values on the Karolinska sleepiness Scale. Studies from Holding et al. (1983) and Hockey et al. (1997) suggest that aversion to invest further effort into task performance is the most reliable characteristic for mental fatigue. This was further highlighted by Boksem et al. (2005), where increased fatigue was associated with a clear decrease in performance on prolonged visual attention tasks.

Table 3.5: Weightings of independent variables on KSS questionnaire using linear mixed model analysis.

Independent variable	Weighting
Time of day	0.183
Core temperature	-1.206
Skin temperature	-0.503
SDNNi	-0.006
Hours into Shift	0.083
HR % above rest	-0.007

No significant differences between pre-post shift for reaction time, minor lapses or major lapses were identified in surgeons using the psychomotor vigilance test, despite a significant increase in self-perceived sleepiness. Whilst it is expected that reaction time increases with time awake, there may be a level of sleep deprivation at the beginning of the shift that is accountable for the non-significant increases in reaction time post-shift, as some surgeons recorded faster scores post-shift. Without monitoring sleep patterns of the night prior, it is difficult to rely on subjective sleepiness alone.

The SURG-TLX has been validated in laboratory settings (Ma et al., 2021) but its use in the field is limited. However, duration of surgery was correlated with complexity and thus increased the mental/physical fatigue, regardless of theatre role. As identified in previous studies, distractions in the operating theatre can affect patient care (Gui et al., 2021). Anxiety was strongly correlated with distractions in the operating theatre in our

study, highlighting the need for opportune professionalism and mitigation of potential distractions in future, in turn reducing anxiety and unnecessary stressors. There was no/near significant difference between duration of surgery and the musculoskeletal soreness, which suggests that surgeons may already be predisposed to musculoskeletal injury prior to the beginning of the operation in question.

Contrary to previous publication (Johnston et al., 2018), the present study identified no correlation for self-perceived sleepiness with control or demand in the workplace. Whilst unexpected, surgeons and nurses shifts offer different demands and tasks. Experience also plays a significant factor for fatigue in surgeons. Older, more experienced surgeons reported more control, greater motor skill capacity, less demanding shifts and less mental effort on their shifts than their younger counterparts. Thus, the duration of surgical expertise seemingly improves the ability to adapt to the demands of the workplace. Whilst sleepiness was homogenous across age groups, a lack of control and high levels of effort are more likely to predispose an individual to fatigue (Hockey, 2011).

Our questionnaire postulates that the surgical population most at risk of becoming fatigued in the workplace are younger surgeons on night shifts resulting from lack of mental stimulation and misaligned circadian rhythms, in line with previous findings (Amirian et al., 2013). Previous studies (Boksem and Tops, 2008; Borragan et al., 2017) have questioned whether prolonged effort induces mental fatigue, or rather under-arousal (e.g. sleepiness/boredom). However in this research, workplace fatigue is identified using the Karolinska sleepiness Scale as sleepier surgeons are the most likely to make errors or near-miss errors, on other tasks outside of surgery. In a real world scenario, it must be considered that participants may not truthfully report errors made due to the credibility of their craft, evaluation apprehension or previously described ‘feelings of invincibility’ (Dorion and Darveau, 2013) which surgeons may experience through rigorous training. Whilst our research identified no errors in the workplace for this study, future research with a larger cohort should be conducted to observe psychological or physiological markers and their associations with errors.

This study highlights the differences in fatigue onset depending on years of surgical experience and shift type; however, there are limitations to the study. Whilst data was obtained from 21 participants (33% of total surgeons), only six were female which is not a true representation of the cohort. Secondly, there is much missing data. Due to the nature of handovers between shifts, surgeons struggled to find the time to wear the equipment before the shift. The protocol did not consider how many consecutive shifts the surgeons had already worked prior to participation, which may have predisposed some surgeons to early fatigue onset when compared to others.

3.3.5 Conclusion

This research identifies novel insight into different shift types and the subsequent demands which the surgeons are subject to. Simultaneously, the protocol is the first of its kind to monitor fatigue levels from a combination of both psychological and physiological markers throughout a surgeon's shift for the identification of workplace fatigue and the causes of its onset. This preliminary study highlights the possibility of fatigue identification with heart rate and skin/core temperature using Linear Mixed Model analysis identifying significant relationships with the Karolinska sleepiness Scale and thus, these wearables are recommended in future studies identifying the onset of fatigue in the workplace.

However, as a limitation, the individual weightings of these variables were small for varying the Karolinska Sleepiness Scale. There was a small sample size which means the results should be interpreted with caution. There was also lots of participants with missing data. A future study may want to use a larger sample size from multiple hospitals, with more technologically advanced equipment that requires an easier set-up to mitigate human error. Additionally, this study identified difficulty using the electrodes. The study was conducted during the COVID-19 pandemic which meant the research team could not be present for the initial data set-up. Though instructions on using the device were administered with the device, there were significant losses of HR data throughout the shift, indicative of improper placement of the electrodes. Equally, feedback from the surgeons in the pilot study had frequent complaints that the

electrodes would pull at the skin when removing them, causing discomfort. Due to the invasiveness of the protocol, the wearable apparatus needs to be replaced.

Whilst the LMM analysis shows promise towards building an algorithm, the LMM does not account for the circadian rhythm variance between the shift types. Equally, there is no normalisation between the different variables, reducing the accuracy and integrity of the data. The data in this pilot study runs on 5-minute epochs, in which much variation can occur when compared to 1-minute epochs. The hours into shift and time of day variables only alter each hour, which reduces the sensitivity of the algorithm. The algorithm does not consider the effects of breaks, which can ultimately reduce fatigue, as they were not monitored in the pilot study. In order to create an algorithm that can be sensitive to the individual, with normalised data as well as fluctuate with breaks, a different approach is required which is based upon comparisons and mathematical cross-correlations, which can individually weight the variables used in the algorithm using chromatic analysis in section 2.4.

Preliminary research on the chosen demographic for this study identified the key contributors to workplace fatigue using a questionnaire, as well as gender differences associated with fatigue, further highlighting the need to objectively monitor fatigue throughout a shift. The pilot study monitored biometrics throughout the shifts of surgeons to identify significant contributors to the KSS, though the weightings of the KSS variance were small and the algorithm does not consider the impact of breaks.

3.4 Summary

This chapter outlines the preliminary research conducted on our cohort in order to collect data in an efficient manner for producing an algorithm for the detection of fatigue in the workplace. Firstly, the surgeon questionnaire is described, which identified an increase in the prevalence of mental exhaustion and musculoskeletal fatigue in females, possibly due to perceived lack of control over workload, high levels of demanding work and prolonged static postures at work. Secondly, the pilot study

results are described. The protocol is the first of its kind to monitor fatigue levels from a combination of both psychological and physiological markers throughout a surgeon's shift for the identification of workplace fatigue and the causes of its onset. The algorithm does not consider the effects of breaks, which can ultimately reduce fatigue, as they were not monitored in the pilot study.

With this information in mind, HR and HRV will be utilised further in the algorithm. Due to the large weighting of the time of day variable, this will also be used in a combined monitoring algorithm, as well as other variables such as accelerometer data and the KSS values that were identified in the pilot study. A different approach is required, other than the LMM, which is based upon comparisons and mathematical cross-correlations, which can individually weight the variables used in the algorithm, using chromatic data processing (section 2.4).

Chapter 4

Experimental methods and test procedures

4.1 Introduction

Once the pilot study and surgeon questionnaire was gathered, a clearer picture was produced as to the effects of workplace fatigue on surgeons and its prevalence. The Linear Mixed Models approach has flaws which are described in section 3.3.5. As such, chromatics will be monitored to predict the onset of workplace fatigue moving forward. This section describes the experimental methods, including all the components of the proposed fatigue identification algorithm. The algorithm contains four biomarkers, Heart Rate (HR), Heart Rate Variability (HRV), Skin Temperature (ST) and Accelerometer data (ACC). Additionally, two sensory models are described; time on shift and circadian rhythm.

The proposed fatigue identification algorithm is designed to predict the onset of workplace fatigue using various components with different weightings. The methodology for the second round of testing is described in detail, as well as how the data is collected, extracted and analysed. Statistical analysis, as well as chromatic involvement, are discussed in this chapter.

4.2.1 Experimental procedure for surgeon fatigue protocol

As with section 3.3.2, participants were recruited through a liaison at a hospital (UK) for data protection. The research team did not have access to any surgeons' details unless they approached the research team voluntarily. Once a participant had registered involvement, the researcher met the participants 15 minutes before their shift to fill out a consent form, set up the equipment and fill out a simple demographic information questionnaire, asking for age range (25-34, 35-44, etc.), gender, department of work and surgical experience (years). After the equipment was worn (described in section 4.4.3), participants were informed they would be sent an automated text message with a hyperlink to a questionnaire to fill out. At the end of the shift, participants returned the chest strap, phone and waist bag to a safe place for the researcher to collect. There were two sets of equipment, so one could be charged and ready when collecting the other set to improve the efficiency of data collection.

In total, 15 participants (eight males and seven females) were recruited for the study. Data was collected between September and October 2022. Participants gave informed consent before participating in the trial. The study was approved by the University of Liverpool's Ethics Committee. The research team analysed data in line with the University of Liverpool's ethical guidance and in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments. Data was only accessed and analysed by the research team. In total, four 24-hour shifts (8 am-8 am), five night shifts (5 pm - 8 am) and five day shifts (8/9 am-5 pm) were analysed. One participant's chest-strap data failed to record and two further participants had poor ECG, therefore their data was excluded from the analysis, however answers from the questionnaire were still analysed. Mean surgical experience (years) was 10.33 (±4.75). Seven participants were aged 35-44, four were aged 25-34 and one was 45-54.

To make further improvements on the pilot study (section 3.3) for the analysis in this chapter, an equal representation of males and females were included in the analysis. The adaptation of the chest strap lowered the artefacts derived from the protocol. The participants had not worked for at least 24 hours prior to their shift to reduce the impact

of work-related fatigue prior to participation in the study. Finally, breaks were monitored as well as the activities on breaks to understand how they affect the overall sleepiness of our cohort.

In order to create an algorithm that prevents the data from being over-fitted, after all participants data were collected and analysed for correlations against the KSS, the participants were divided at random into two groups using a random number generator for chromatic analysis. The first group were refined and analysed so that the strongest correlation could be achieved with those participants as a proxy for identification of KSS without using the questionnaire. The algorithm was then fitted with the second group of participants, to see whether the algorithm was accurate when tested again. More information on the refinement of the methodology, can be found in section 4.3.

4.2.2.1 Surgeon questionnaire

In order to make improvements on the pilot study, the questionnaire in section 3.3.2 has been modified. The questionnaire was sent by automated text every three hours which contained a hyperlink to a webpage. To see how breaks affect the sleepiness of our participants, we asked the following questions every three hours;

1. How sleepy do you feel?
2. Have you made an error or near-miss error since the last questionnaire?
3. Have you had a break yet?
4. What time did your break start/finish?
5. Did you consume caffeine, nap or smoke on your break?
6. How sleepy were you at the beginning of your break?
7. How sleepy were you at the end of your break?

At the beginning of each shift, the demographic questionnaire also asked how long the participants had slept for prior to their shift, as well as the quality of their sleep (good, average or poor).

Participants were asked whether they consume caffeine or smoke as they are stimulants which effect Heart Rate and also subjective levels of sleepiness (Wetter and Young, 1994; O'Callaghan et al., 2018) . Naps have been shown to reduce tension and suppress the sympathetic nervous system at work (Oriyama et al., 2014). Monitoring these variables may show trends towards decreased subjective tiredness on shift, compared to those who do not.

4.2.2.2 Karolinska Sleepiness Scale as a measure of Fatigue Index

The Karolinska Sleepiness Scale (KSS) ranges from one (extremely alert) to nine (very sleepy) (Shahid et al., 2011). Is has been used extensively in the literature to identify fatigue and sleepiness in both laboratory and field settings (Hirshkowitz and Sharafkhaneh, 2017). For this study, the participants answered the KSS every three hours on shift, as to how they feel when answering the questionnaire. Data were linearly interpolated between questionnaires, for example, a participant answered a '3' at 9am and a '5' at 1pm. The data will be linearly extrapolated between the data points in Microsoft Excel (section 4.4.2) in equal epochs (e.g. every 5 minutes/ every row).

For this study, any answers '$\geq$six' on the KSS were deemed as a cause of concern, as this point on the KSS is described as 'some signs of sleepiness', suggesting the need to take a break from work to mitigate the chances of causing errors in the workplace.

4.2.3 Chest strap

To improve the protocol after feedback from the pilot study in section 3.3, a chest strap was used for the second round of testing with the participants. The chest strap provides added comfort and less artefacts than the previously used electrodes. The chest strap measures ECG trace and collects data at 320Hz. Additionally, the chest strap measures ST at 1Hz and ACC at 24Hz. The chest strap consists of a magnetic pod, which connects to the strap. The pod contains Bluetooth capabilities. Unlike the pilot study (section 3.3), the research team were able to hand over the equipment in person, to reduce the likelihood of potential errors when administering the equipment. Participants were

asked to wear the strap so that the pod sits underneath the left arm for proper alignment of the ECG sensors.

4.2.4 Mobile application for data collection

For the data to be extracted from the chest strap, data had to be downloaded on to a mobile phone via Bluetooth. As such, the chest strap had to stay within 3 meters of the mobile phone at all times. There was no interference caused by the system or surgical equipment. The mobile phone was the model TTfone TT20, used solely to contain an application provided by the manufacturer of the chest strap for data extraction purposes. The mobile phone was provided with a waist bag, so the participants did not accidentally leave the phone anywhere, thus disconnecting the Bluetooth connection. The application (Image 4.1) allowed the user to allocate a length of time in seconds of each file to be downloaded, and the total duration the application connects to the pod. For the purposes of this research, a duration of 900 seconds was chosen for each data file. For example, a participant working a 12-hour shift would save in 900-second intervals for a total duration of 43,200 seconds (3600 x 12). The application allows the user to view their ECG trace before the collection of data, to ensure the sensor is lined up correctly. After data collection, the phone can be connected to a computer via Micro-USB for data analysis.

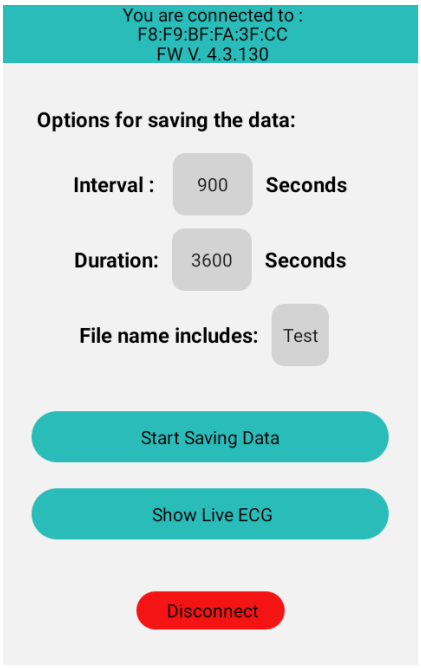


Image 4.1: Screenshot of display screen for mobile application used to collect data, showing four data files saved in intervals of 900 seconds.

4.3 Variables integrated into composite system.

The following section includes the variables that are to be analysed for possible algorithm integration. The variables proposed are four physiological measures (components a-d) and three physiological models (components e-g):-

a. Heart Rate (HR)
b. Heart Rate Variability (HRV)
c. Skin Temperature (ST)
d. Accelerometer data (ACC)
e. Circadian rhythm
f. Time on Shift
g. Sleep Quality

4.3.1 Proposed model for identification of workplace fatigue

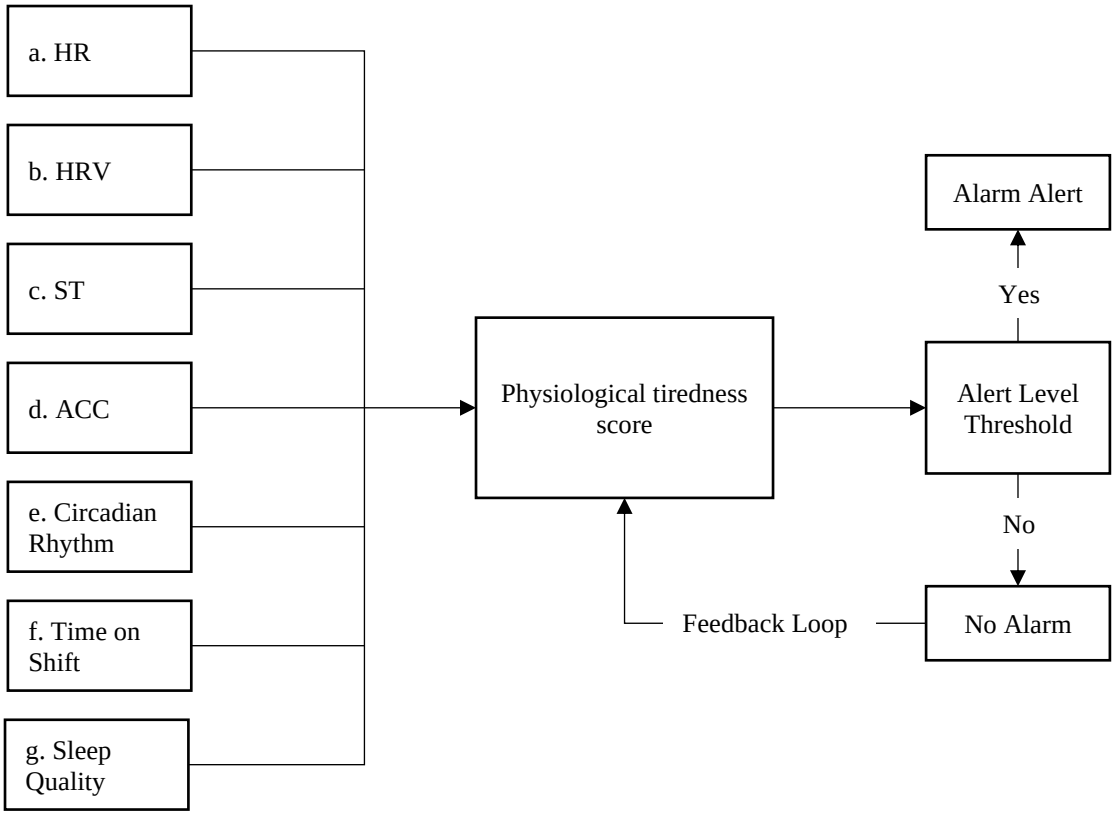


Figure 4.1: Proposed model for identification of workplace fatigue advisory system, consisting of seven components a-g.

Figure 4.1 (above) shows the proposed components for identification of workplace fatigue based on the preliminary research outlined from the pilot study within the capabilities of the new and improved wearable technology. The model shows, that the components (a-g) are combined to produce a physiological tiredness score, whereby exceeding a set threshold prompts an alarm to trigger. The weighting and inclusion of components a-g is to be researched.

The components of the algorithm are normalised between min and max parameters for individual weightings depending on the relationship when correlated against the KSS. Quantification of the min and max parameters can be found in section 4.5.5 and the weightings of each variable are found determined using correlation analysis, reported in section 4.4.2. The following section explains how each variable is quantified in the algorithm. The proposed fatigue index algorithm is displayed in section 4.3.2.

a. Heart Rate (HR)

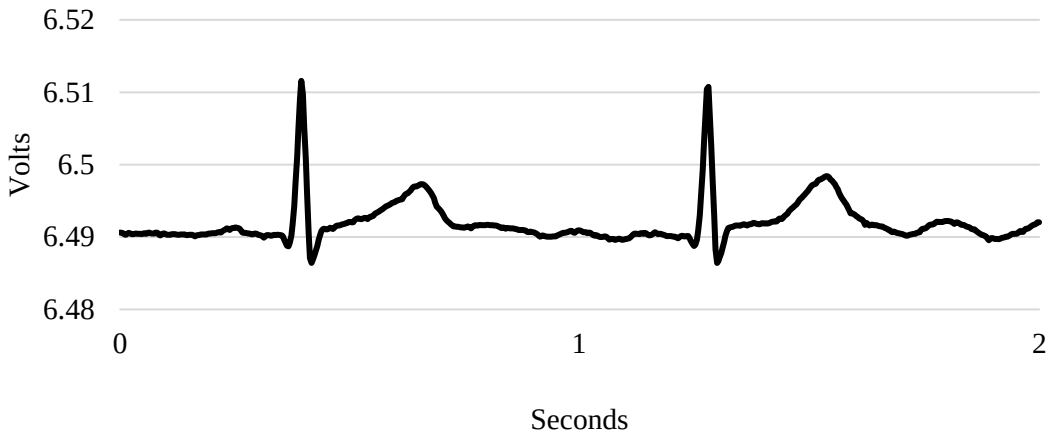


Figure 4.2: Heart Rate (displayed as electrocardiogram) versus Time.

The first component of the proposed model is HR (Figure 4.1), which is unique to every individual with reference to electrocardiogram trace (ECG trace), minimum and maximum heart rate. An example of 2 seconds ECG trace is given on Figure 4.2, containing 640 samples.

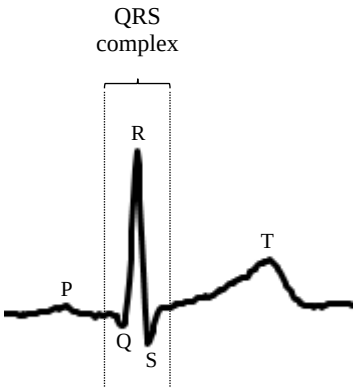


Figure 4.3: ECG trace defining key features, e.g. QRS complex, for one heartbeat.

Figure 4.3 shows the breakdown of the ECG trace for one heartbeat. Figure 4.4 is an example of one participant's HR throughout their shift. Their HR is calculated from the time in between beats, also known as the 'R-R Interval' (RRI), which is extracted from the mobile app (section 4.2.4). The RRI is taken from the large spikes on each heartbeat. On the ECG trace, these spikes or 'peaks' are known as the 'R peaks'. Figure 4.4 contains the average HR for every minute on shift. The graph clearly shows the fluctuation of HR throughout the shift, with varying levels of intensity. There are periods where the HR is relatively stable and there are HR spikes during the shift, when the participant is active or carrying out stressful work.

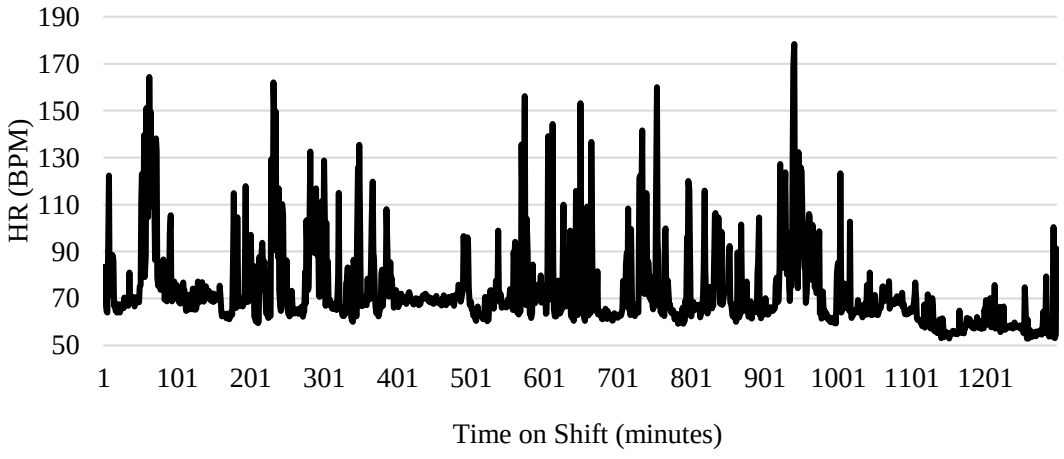


Figure 4.4: One participants heart rate throughout their shift versus time.

b. Heart Rate Variability (HRV)

The second component is a more in-depth analysis of HR. As mentioned before, HRV can be quantified in the time or frequency domain, both of which are examined in this research. For the purposes of this analysis, the time domain measures will be utilised due to their lower computational cost. Previous research in section 3.3.2 utilised the standard deviation of N-N intervals (SDNN), quite simply the standard deviation between heartbeats in milliseconds (ms). SDNN is the gold standard for stratification of cardiac risk over a 24-hour period, but less effective during shorter periods as they are not as representative of environmental stimulation [Shaffer and Ginsberg, 2017]. However, upon further research the root mean square of successive differences (RMSSD) has been used extensively for short-term ECG analysis. As the data for analysis in this study uses 15-minute intervals, RMSSD is a more acceptable measure of stress in the workplace.

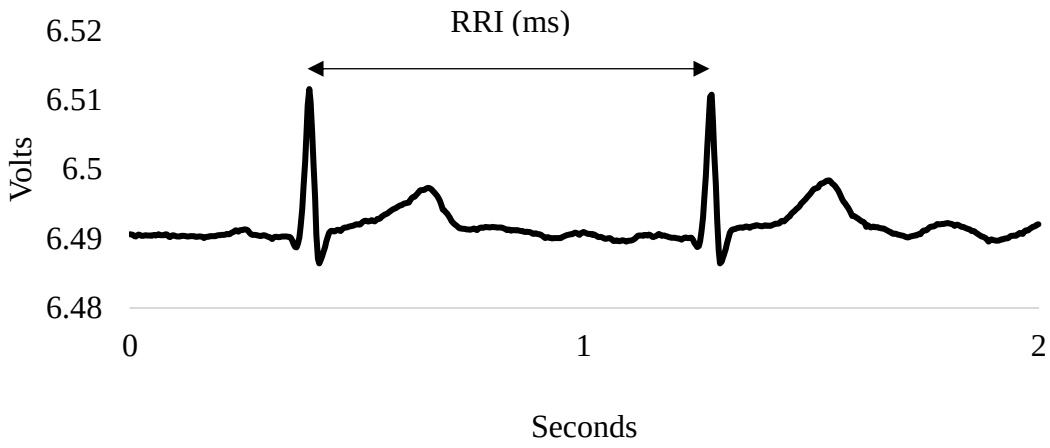


Figure 4.5: Illustration of heart rate variability (figure 4.2) with added demonstration for the measure of RMSSD using ‘R-R interval’ versus time.

RMSSD reflects the RRI variance in HR. The RRI (R-R interval) measures the time elapsed between two successive ‘R’ peaks on the QRS signal of the electrocardiogram (Figure 4.3) RMSSD is inversely related to the activity of the sympathetic nervous system (SNS); higher RMSSD values indicate higher parasympathetic dominance and lower SNS influence on HRV, while lower RMSSD values suggest increased sympathetic activity and reduced parasympathetic influence. RMSSD is obtained by calculating the successive RRI’s in ms. Then, each value is squared and the result is

averaged before the square root of the total is obtained. The conventional minimum recording is 5 minutes, however ultra-short recordings of 60-seconds have been proposed [Salahuddin et al., 2007, Nussinovitch et al., 2011, Baek et al., 2015], as long as artefacts are monitored and carefully removed. The SNS prepares the body for physical activity, through redirection of oxygenated blood around the body in response to physical or mental stimuli.

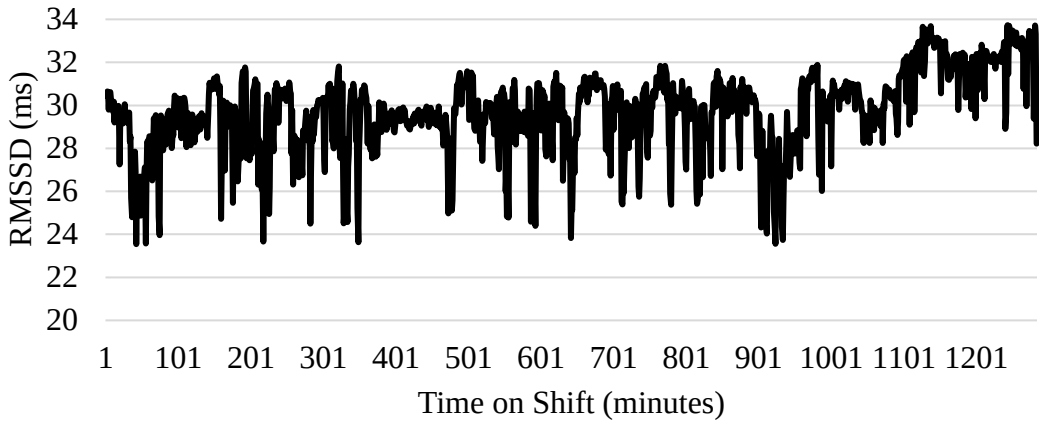


Figure 4.6: Example of RMSSD for one participant versus time.

Figure 4.6 is an example of RMSSD from the same participant as Figure 4.4. The RMSSD measure could be interpreted as inversely proportional to HR, whereby increased HR causes decreased RMSSD measures.

Both components; HR and HRV, can be affected by multiple characteristics, such as age and sex. Age causes HRV measures to decline with age, maximal HR also declines with age and a common estimate for maximal age-related fatigue is, 220 – age. Females have a higher mean HR (and smaller RRI) when compared to males. Therefore, the proposed model will consider these variables and the effect they may have on fatigue.

c. Skin Temperature (ST)

The skin is the human body's primary cooling system and as such deviations in temperature can increase the stressors on the body to induce fatigue [Or and Duffy, 2007]. Though less reported than HR and HRV, ST has been researched as a potential indicator of fatigue. As fatigue, stress or anxiety increases, this can cause an increase in sweating due to increased SNS activity. The monitoring of sweating is called

electrodermal activity (EDA). EDA measures the change in skin resistance associated with sweating, therefore it is a sensitivity index of the SNS. Unfortunately there is currently no technology the research team could acquire to measure and extract data for EDA in the workplace. However, the research team were able to measure the temperature of the skin, with conductive sensors on the chest strap. Measuring temperature on the torso is less susceptible to environmental changes, such as on the forearm, for example when walking outside. Figure 4.7 shows an example of ST throughout a shift. ST can be altered by environmental constructs, as well as pressure on the sensor. Nonetheless, information may be derived from the temperature sensor that may indicate the onset of fatigue in the workplace, for example long periods of high or low ST.

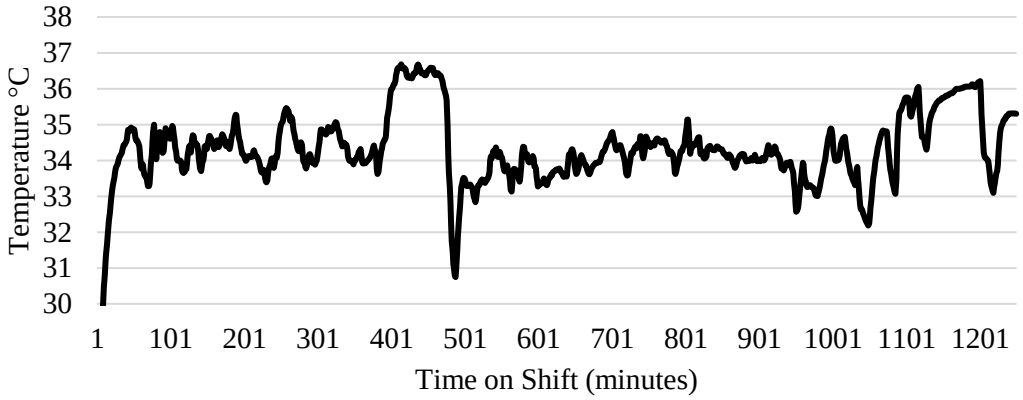


Figure 4.7: Illustration Skin Temperature for one participant on shift versus time.

d. Accelerometer data (ACC)

Accelerometer data (ACC) measures the acceleration of a moving or vibrating body. Generally, ACC are triaxial, providing simultaneous measurements in three orthogonal directions. These directions are typically referred to as *X, Y* and *Z*. Quick movements produce large spikes of ACC, whereas periods of time with no or little movement produce small data outputs. Accelerometers can be placed anywhere on the body, usually also consisting of gyroscopes or magnetometers to produce Inertial Measurement Units (IMUs). However, for this research only the ACC is provided for analysis. Individual manufacturers have their own algorithm for activity classification, which is obtained from accelerometer data. Therefore, there is no single algorithm used

for the detection of activity classification. Activity classification could be listed as; walking, resting, exercise, falling, etc.

HRV and activity recognition in unison are more promising at identifying chronic cognitive and physical fatigue levels than the two variables individually (Gonzalez et al., 2017), showing promise that a combination of variables may be a more robust predictor of fatigue onset. Surgical staff may spend much of time standing or performing activities that require some physical effort, therefore they are at a greater risk of depleting energy sources and becoming physically fatigued when compared to more stationary counterparts (Johnston et al., 2018). However this study is not looking at physical fatigue which is more associated with athletes. This study will want to identify the onset of workplace fatigue, a multidisciplinary construct consisting of physiological and psychological constructs. Therefore, ACC cannot predict workplace fatigue alone, however there may be identifiable trends which increase the likelihood of fatigue onset. Figure 4.8 shows the total change in magnitude of acceleration in *x, y* and *z* direction for each sample, averaged for every minute (vector magnitude = $\sqrt{(X^2 + Y^2 + Z^2)}$. This calculation allows a visual representation of when the participants are active or stationary on shift. Large spikes in the figure can be categorised as high levels of activity, for example walking, whereas stable levels can be categorised as low levels of activity, when sitting or lying down.

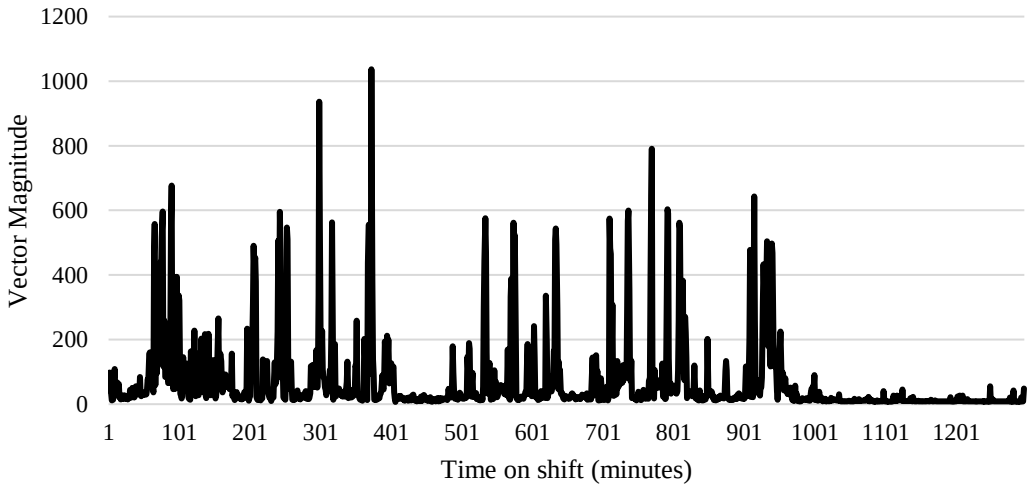


Figure 4.8: Illustration of sum of accelerometer vectors for one participant on shift.

e. KSS model (Circadian Rhythm)

As mentioned in section 2.2.2, Karolinska sleepiness scale (KSS) and circadian varies with the time of day and is generally unique to the individual. It can be misaligned by lack of sleep, jet lag or working night shifts. Based on the information derived from section 3.3 (explained in section 2.2.2), Figure 4.9 indicates the relevant Karolinska sleepiness scale predictor for the participants in the study used in the proposed model (roughly inversely proportional to circadian rhythm (adopted from Bowman et al., 2002)). The figure represents the predicted average sleepiness, based on the KSS, from the surgeons in the pilot study. Day and 24-hour shifts increase over time, however the night shift has a unique pattern beginning high and steeply declining, and declining again in the early hours of the morning, which may be due to the suppression of melatonin with increased natural light.

The values depicted in the graph are stationary and identical for all participants. For example, if a day shift (08:00:00am-20:00:00pm) participant's data collection began at 9:00:00am instead of 08:00:00am, the circadian graph would line up accordingly.

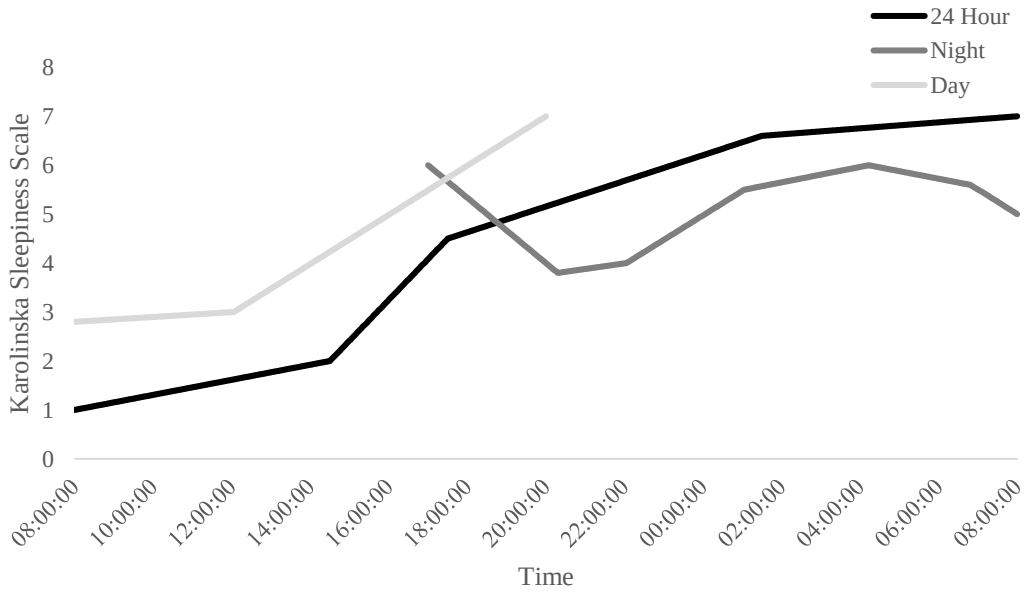


Figure 4.9: Karolinska sleepiness scale model for the three shift types, day, night and 24-hour (roughly inverse to circadian rhythm model (Figure 2.3)).

f. Time on Shift

Time on Shift accounts for the inevitability of time awake – the increase in tiredness as the shift progresses (Johnston et al., 2018). Figure 4.10 shows the increase in score added per minute, depending on the length of the shift. This increase in score is adopted from Koh et al. (2007). The rise in component f. is consistent for all participants. Out of all the components in the proposed model, this component fluctuates with the implementation of breaks, which are included in the shifts. To delay the rise in tiredness depicted by the Time on Shift component, breaks and their effect of reducing the KSS are considered. This increase is derived from previous research into driver fatigue (Reyner and Horne, 1998), as there are currently no figures to support the rise in workplace fatigue. Table 4.1 below shows the rise in score every minute, which increases each hour into the shift. The score increases by 0.00125 for each hour into the shift. For example, if a participant has been on shift for 6 hours, then the score would increase by (0.00125 x 6) per minute. The Time on Shift variable is shown in Figure 4.10, which is the same for all participants.

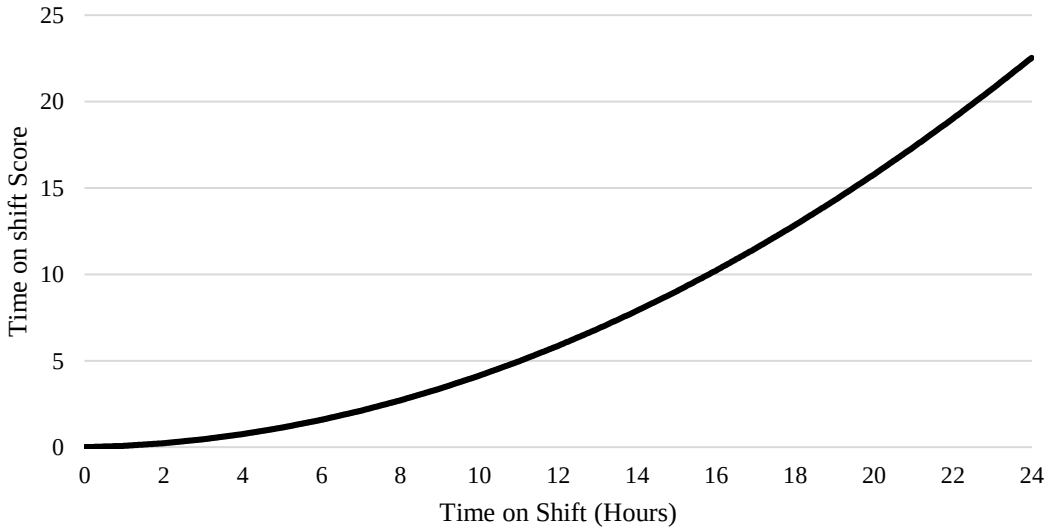


Figure 4.10: Time on Shift score as a function of time, with increased steepness over time.

Table 4.1: Increase in score of Time on shift variable per minute versus hours into shift

Time into Shift (Hours)	Increase in score (per minute)
0	0.00125
4	0.005
8	0.01
12	0.015
16	0.02
20	0.025
24	0.03

g. Sleep Quality

The sleep quality component is dependent on the participants answer to their 'quality of sleep in the pre-shift questionnaire (section 3.2.2). This methodology proposes that sleep quality effects the Time on Shift (component f), whereby individuals who score 'poor' or 'average' sleep have a shift from baseline of the Time on Shift component on the y-axis (Figure 4.10). However, the results from the pre-shift questionnaire and the resultant peak KSS scores from the individuals will test this hypothesis and calculate the respective shift in the Time on Shift variable.

4.3.2 Fatigue Index for advisory system

A proposed mathematical algorithm (Eqn 7, below) to calculate the fatigue index respective of the Karolinska Sleepiness Scale (KSS) values for the new cohort of surgeons combines the seven components aforementioned (section 3.3.1), adopted from Koh et al. (2007). Each variable is to be weighted according to their influence on the KSS using correlation analysis (section 4.4.5), using Eqn 7. Further testing will determine the weighting and the inclusion of the variables.

$$FI = \Sigma(HR(t) + HRV(t) + ST(t) + ACC(t) + KSS\_Pred(t) + (ToS_{SQ}(t) - DR(t))) \quad \text{--- Eqn 7}$$

Where,

- FI = Fatigue Index
- HR(t) = Heart Rate with respect to time
- HRV(t) = Heart Rate Variability with respect to time
- ST(t) = Skin Temperature with respect to time
- ACC(t) = Accelerometer data with respect to time
- KSS_Pred(t) = Predicted KSS with respect to time
- $ToS_{SQ}(t)$ = Time on Shift with respect to time and sleep quality
- DR(t) = Duration of rest with respect to time

4.4.1 Data analysis

The data extraction for the questionnaires and mobile phone are described in this section. There were very large amounts of data to work with in this protocol. As such, platforms MATLAB (MATLAB ver. R2020a) and IBM SPSS Statistics (Version 27) were used to compress and analyse the data.

4.4.2 Biometrics extraction from mobile phone

HR, HRV, ST and ACC were extracted from the mobile device using Microsoft Excel and MATLAB. All biomarkers are initially saved in one Excel document, in 900-second intervals. Conversion of the raw data into a computable format was conducted using an APK file (file format for applications used on Android operating systems) provided by the manufacturer. Additionally, the manufacturer provided a .exe file which converted the ECG channel into an RRI format. The programmes listed are described below.

1. **Microsoft Excel**

Microsoft Excel (Microsoft Corporation, 2016) is a spreadsheet programme, which allows the free manipulation of large data sets. Formulae and multiple spreadsheets can be utilised in order to provide meaningful information within strategically ordered columns and rows for further analysis. An example of the spreadsheet is given in Appendix C.

2. **MATLAB**

MATLAB (MATLAB ver. R2020a) is a programming platform designed for engineers to analyse and design systems through code, making complex tasks much easier using customisable scripts. Computational mathematics can be saved and adjusted according to the needs of the user, without repeating the code with each use. An example of MATLAB code used in the project is given in Appendix C.

4.4.3 Technical specification of chest strap

The performance characteristics of the chest strap can be found in Appendix C. The chest strap pod has a rechargeable lithium polymer battery, with a minimum battery life of 24 hours, meaning it lasts for the duration of the 24-hour shifts some surgeons in our study worked. Multiple chest straps were provided for each size, ranging from extra-small to extra-extra-large.

4.4.4.1 Extraction of HR and HRV data

4.4.4.1.1 Time domain

Data for HR (Figure 4.4) requires the manufacturer's algorithm in order to reconstruct the packets of raw data and reorganise them into time series per channel. The black-box algorithm provided by the manufacturer exports all the raw data and saves them into new .csv files (comma-separated values), three-channel ECG and RRI peaks, respectively. The three ECG channels produce three RRI peak columns, as there are disparities between the ECG channels (examples of these files can be found in Appendix C). For each RRI peak column (RPeak_Ch1, RPeak_Ch2 and RPeak_Ch3), a ratio is also saved in the adjacent excel column which compares the current R-peak

vs baseline to the next R-peak vs baseline (an example is in Appendix C). The higher the ratio, the better the result, as it indicates that the baseline between QRS complexes in the ECG trace do not have high peaks from noise or motion artefacts. However, this may be biased if the R-peak itself is a false positive from a QRS-like artefact.

As such, to minimise the artefacts presented in the HR and HRV, the standard deviation (SD) of the HR calculated from the three RRI peak columns were compared against one another. The column with the lowest SD was used for final analysis for that 15-minute segment of the overall HR analysis, as the files were saved in 900 second intervals.

In order to calculate the HR from the RRI peak columns, MATLAB was used. The RRI peak columns show a cumulative increase in Hz over time. The HR is sampled at 320 Hz. The code used in MATLAB for the extraction of HR and HRV can be found in Appendix C. For example, the first five heartbeats from the sample are displayed as:

263, 475, 686, 902, 1121…

To calculate the HR from these figures, the respective RRI peak columns are saved as .txt files in the Microsoft application 'Notepad'. Then, they are loaded into MATLAB and the code quantifies the HR as shown:

90.566, 90.995, 88.888, 87.671…

Thus, smaller RRI peak column differences quantify faster HR. For the extraction of our HRV measure, RMSSD, the HR needs to be converted into milliseconds (ms), which is the time-domain measure of HRV:

662.500, 659.375, 675.000, 684.375…

However, the analysis conducted for this research monitors the average HR and HRV measures every minute. Therefore, the total number of heart-beats is divided by 15 to calculate the average per minute. There are instances however, where the length of BPM is not divisible by 15, and the residual heartbeats are stored in a 16th block. These are removed from analysis due to the small proportion of leftover heart beats which are not divisible by 15. The variable 'U' must be rounded as the built-in MATLAB variable 'blockproc' cannot compute fractional parameters (code found in Appendix C).

For example, a 15-minute sample may contain 1369 samples:

1369/15 = 91.267 → 91 when rounded.

91 x 15 = 1365 → 4 heartbeats are leftover and create the 16th variable.

4.4.4.1.2 Frequency Domain

MATLAB is able to compute the power spectral density (PSD) from the 15-minute ECG samples. For each 15 minute epoch throughout the participants shifts, the raw ECG data is loaded from a text file, and the sampling frequency is defined as 320 Hz. The raw ECG signal is then plotted to visualize the data (Figure 4.11). To remove noise, spectral leakage and baseline wander, a high-pass filter is applied to the ECG data in order to remove frequency components below a specified cut-off frequency of 0.04Hz (Figure 4.12a). A second order Butterworth is used to design the filter co-efficients (see Appendix D). MATLAB's 'filtfilt' function filters the signal forward and backward to eliminate phase distortion, resulting in a filtered ECG signal with frequencies below 0.04Hz removed. The filtered ECG signal is then differentiated, squared, and integrated

to obtain a tachogram, which is a plot of the time differences between consecutive R-peaks in the ECG signal (Figure 4.13).

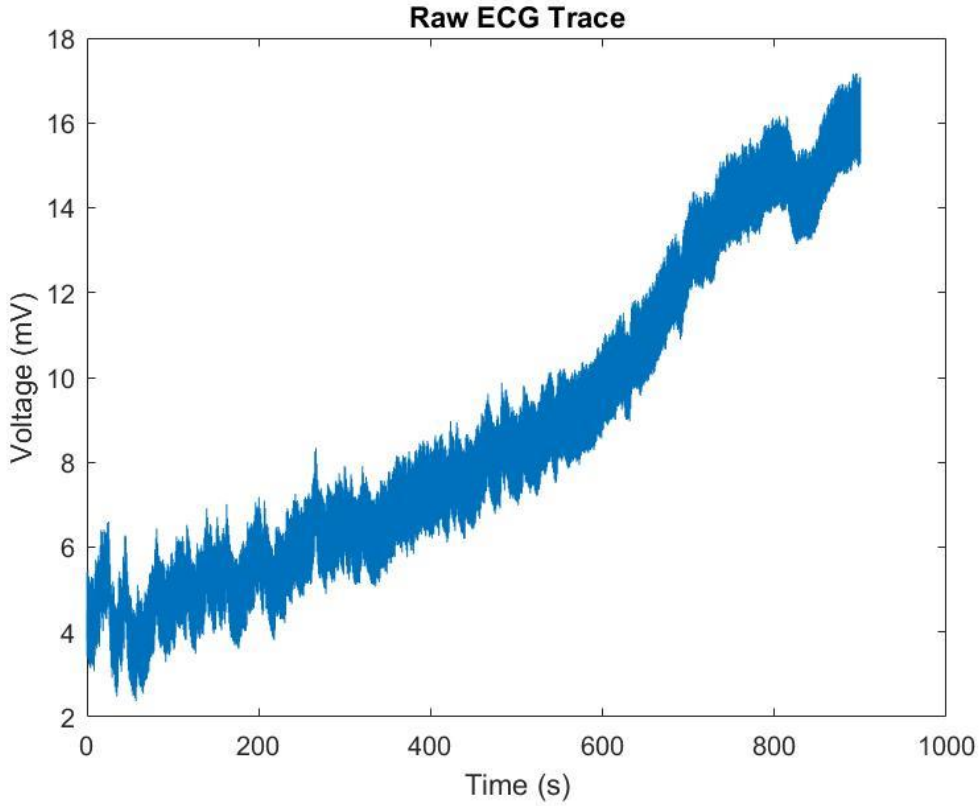


Figure 4.11: Raw ECG trace showing baseline drift and noise.

The R-peaks are identified using the Hamilton-Tompkins algorithm, and the tachogram is divided into three equal sections, 5-minute samples. Five minute samples were chosen as they are the minimum measurement time for the analysis of frequency domain measures (Shaffer and Ginsberg, 2017). For each section, the Lomb-Scargle periodogram is computed to determine the power spectral density of the tachogram. The power spectral density is calculated for three frequency bands of interest: ultra-low frequency (ULF) with a range of 0 to 0.04 Hz, low frequency (LF) with a range of 0.04 to 0.15 Hz, and high frequency (HF) with a range of 0.15 to 0.4 Hz. The indices of the frequency bands are found using the find function in MATLAB. The ultra-low frequencies (ULF) (0.003-0.04Hz) are not specifically associated with stress or fatigue, typically referring to thermoregulatory processes or slow physiological mechanisms (Shaffer and Ginsberg, 2017).

a) b)

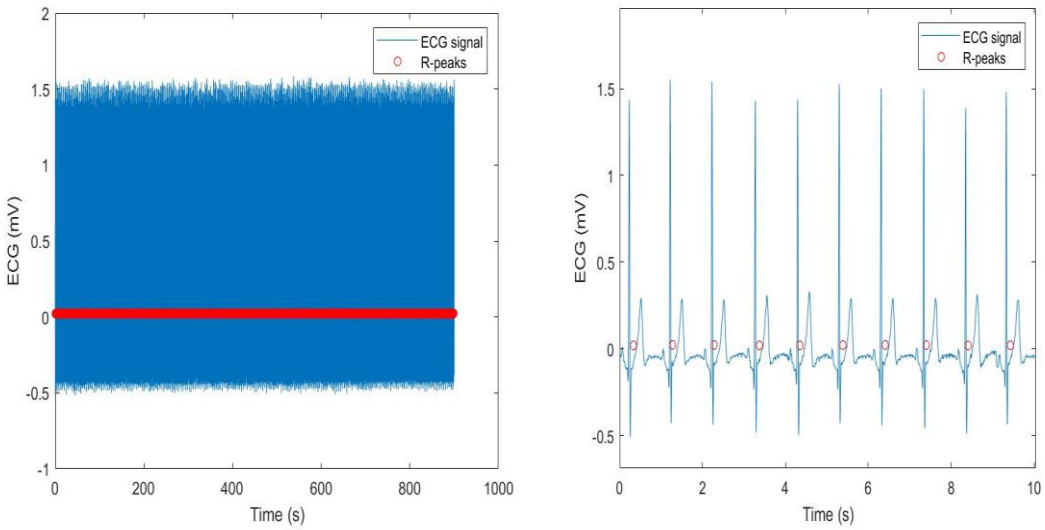


Figure 4.12a and 4.12b: a) Filtered ECG trace showing removal of baseline drift and noise whilst plotting individual R-peaks, and b) a 10 second snapshot showing the r-peaks, outlined as red circles.

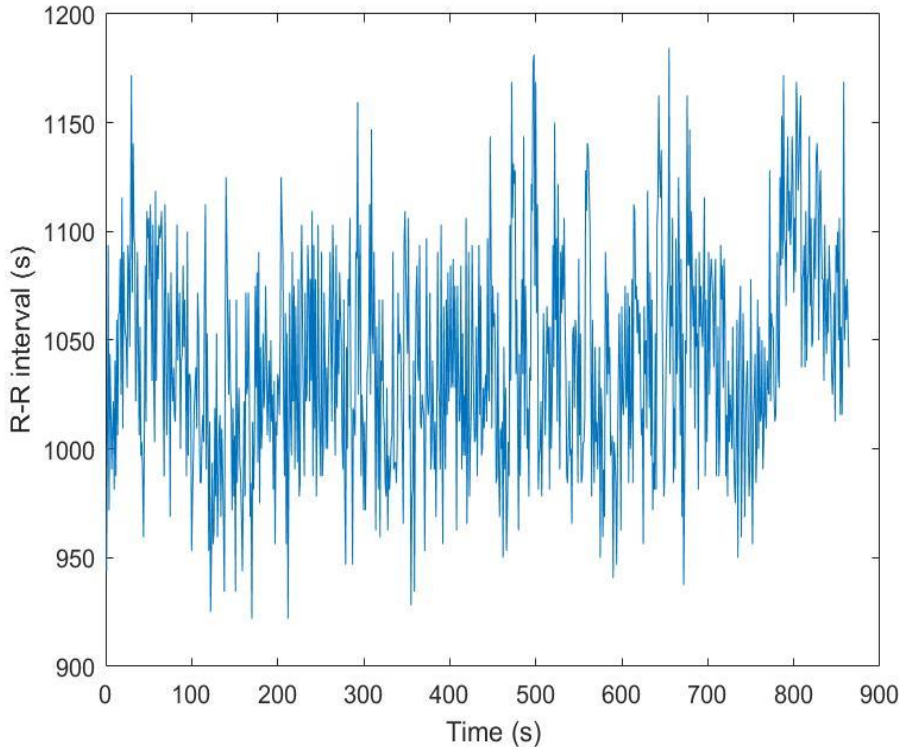


Figure 4.13: Tachogram showing the time in between heart beats during 15-minute sample.

For each section, the LF and HF powers are calculated by summing the power spectral density within the respective frequency bands. The LF/HF ratio is calculated as the ratio of the LF power to the HF power. The mean power spectral density is computed for each section using the mean function in MATLAB. All values for LF power, HF power, LF/HF ratio, and mean power spectral density are stored in separate arrays for further analysis. As shown in Figures 4.16 and 4.17, the x-axis ranges from 0.04 – 0.4 Hz, removing the ULF components of the ECG signal, which are not involved with stress or fatigue.

The MATLAB code also includes plotting functions to visualize the results. Specifically, the power spectral density is plotted with frequency on the x-axis and power on the y-axis. The LF and HF frequency bands are highlighted in red and blue, respectively. The area function is used to create shaded regions for the LF and HF frequency bands. The plot also includes a legend and labels for the x and y axes (Figure 4.14).

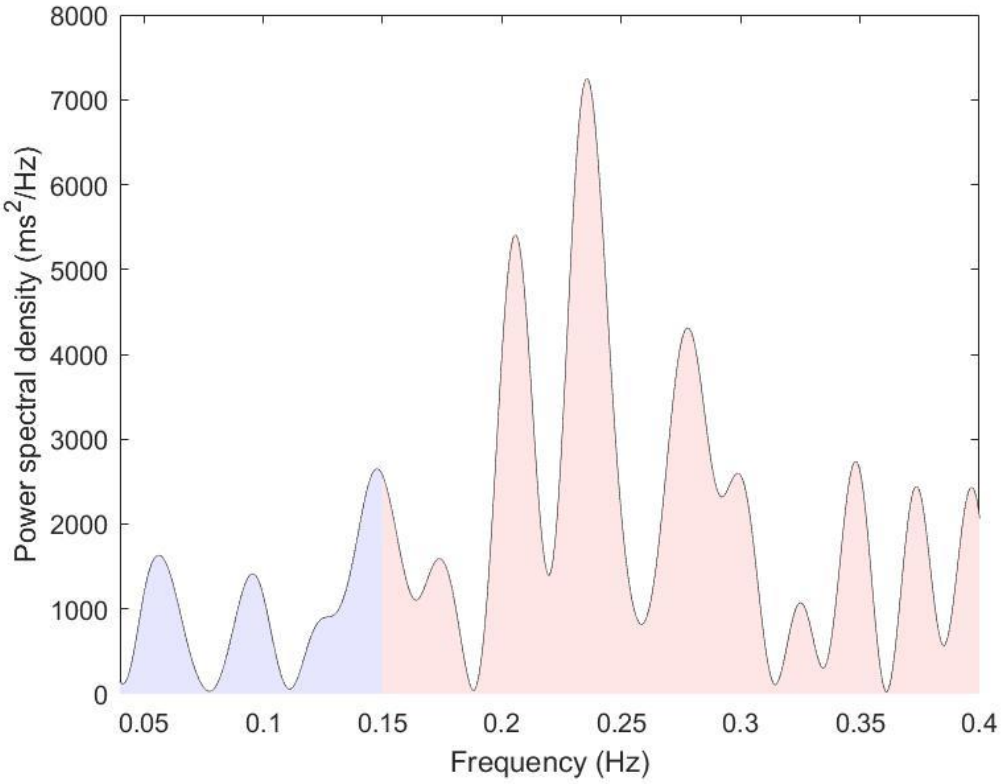


Figure 4.14: The power spectral density (PSD) of one 5-minute sample of filtered ECG data for the analysis of frequency domain measures.

Finally, the chromatic filters R, G and B are superimposed over the PSD charts, for analysis of the dominant filter. Figure 4.15 shows the chosen superimposed filters on top of an example of PSD. For the analysis of PSD, the chosen filters are truncated triangular filters with a shortened green filter. The truncated triangular filter (Koh et al., 2005) is used for the processing of discrete data sets, where one component of the data is independent from the next, whereas triangular or gaussian filters are more appropriate for continuous data sets. The shortened green filter was chosen to further discriminate the areas of low frequencies and high frequencies, (R and B filters, respectively), for the analysis of low or high frequency dominant signals. Whereby, low frequency dominance results in a dominant R filter (sympathetic nervous system dominance) and a dominant G or B filter results in high frequency dominance (parasympathetic nervous system dominance). Justification for this filter choice in Figure 4.15 can be found in section 4.5.2.

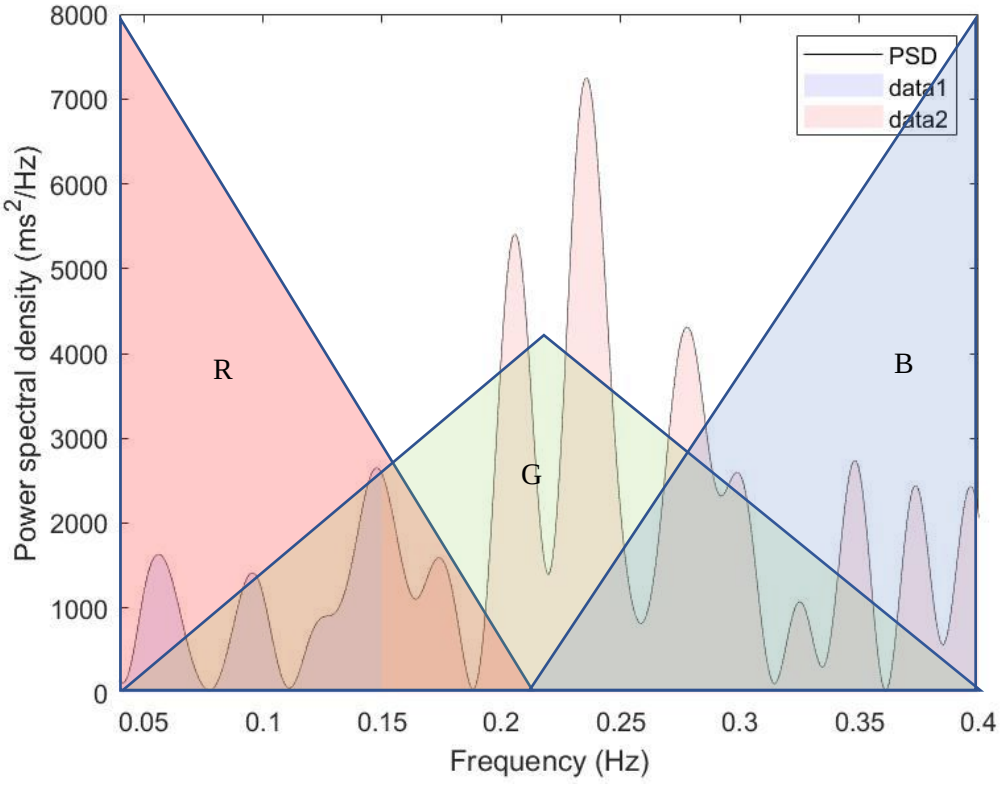


Figure 4.15: The power spectral density (PSD) of one 5-minute sample with modified superimposed, truncated chromatic filters (filters are labelled, R = red, G = green and B = blue.

Using this novel approach to the chromatic filters allows discrete distinction between each five-minute sample of ECG. This allows the research team to outline periods of stress or rest with a minimum of five minutes. The R and B filters meet at 0.022Hz, the R filters exclusively looks at the low frequency components of the PSD (0.04-0.15). In the example shown in figure 4.15 the dominant filter is G, which demonstrates high frequency dominance. The LF/HF ratio in this figure is 0.19, indicating greater parasympathetic activity and lower sympathetic activity (Area of LF divided by area of HF). The PSD can be used to identify 5 minute periods at work and assess, in greater detail than the RMSSD alone, the level of stress upon the individual during the previous five minutes of work.

4.4.4.2 Extraction of ST

After the manufacturers black-box algorithm separates the 15-minute files into the relevant biomarkers, new .csv files are constructed which contain the pod temperature in 1Hz samples. Each new file contains 900 samples of data. Each sample of data is Unix timestamped. The data is saved into a .txt file and loaded into MATLAB to find the average of every minute's worth of data.

4.4.4.3 Extraction of ACC

Similarly to ST and HR extraction, new .csv files are constructed containing a timestamp and tri-axial accelerometer, namely *x, y* and *z*. The raw accelerometer output for *X, Y* and *Z* are combined in a process called vector magnitude processing (VM). VM thus represents the intensity of acceleration the device is subjected to, regardless of direction. No data preparation (e.g. calibration, conversion into g force) was conducted in this experiment. An example of the raw data, which is collected at 30Hz, is outlined in Table 4.2 below.

Table 4.2: Example of Vector Magnitude raw data calculated for accelerometer analysis

Timestamp (sec)	*X*	*Y*	*Z*
0	-726	4046	-1127
0.033	-670	4061	-1189

In order to calculate the resultant VM, the respective squared values of X, Y and Z columns are calculated for each time stamp and then summed. The total sum is square rooted to provide the value *T*, as shown:

$$\sqrt{(X^2 + Y^2 + Z^2)} = T$$

Then, the most recent timestamp is taken away from previous timestamp, to calculate the change in vector acceleration. Using Table 4.2, the equations are shown below. For instances where the delta produces a minus number, this is multiplied by -1 to produce a positive integer:

$$\sqrt{(-726^2 + 4046^2 + -1127^2)} = 4136.8$$

$$\sqrt{(-670^2 + 4061^2 + -1189^2)} = 4284.2$$

$$(4284.2 - 4136) = 148.2$$

Finally, the delta between each timestamp is saved into a .txt file for analysis in MATLAB, to calculate the average acceleration per minute, resulting in a graph similar to Figure 4.8.

4.4.4.4 Normalisation of parameters used in algorithm

As mentioned in Eqn 1 (section 2.4.2), the use of chromatics requires the normalisation of different signals on to a common range 0 to 1 (Zhang, 2004). However, Eqn 1 requires the maximal amplitude of the signal for the entire duration, which cannot be calculated whilst collecting data in real time. As such, minimum and maximum parameter sizes have been pre-determined through the collection of data (section 4.2.1) in order to provide a threshold of normalisation. The newly formed equation (Eqn 8) is below:-

$$X_{normalised} = \frac{(X - \min(X))}{(\max(X) - \min(X))} \qquad \text{---------- Eqn 8}$$

Eqn 8 enables the chromatic output to yield real time normalisation of a signal (X), such as the parameters for HR mentioned in section 4.5.5.

Table 4.3 outlines the minimum and maximum parameters used for each variable, respective of shift type. The table depicts the 8am-5pm day shift. All parameters are the same for Day shift starting at 9am-5pm, apart from the maximum Time on Shift parameter is 3.388 (Table 4.3). The value in Table 4.3 are determined by the respective time points from Figure 4.10 and Figure 4.9, respective of shift type.

Table 4.3 Minimum and maximum parameters for all variables in algorithm

Variables	Day shift (8am-5pm)		Night shift		24 hour shift	
	Min	Max	Min	Max	Min	Max
RMSSD	25	30	25	30	25	30
Predicted KSS	2.8	7	3.81	7	1	7
Time on Shift	0	4.1388 3.388*	0	9	0	22.4275

*9am – 5pm

4.4.5 Statistical analysis

SPSS (statistics 28) was used for statistical analysis. To evaluate the relationship between the proposed variables and KSS readings in surgeons, Pearson correlation coefficient analysis and regression analysis are performed (Freedman et al., 2007, Freedman et al., 2009). T-tests and ANOVAs are used to show the mean differences between gender and shift-type, respectively (Thomas et al., 2011). Where parametric assumptions are not met, non-parametric tests are used for t-tests and ANOVAs (Moore and McCabe, 2004). The non-parametric tests are Mann-Whitney U an Kruskal-Wallis tests (McKnight and Majab, 2010; Moore and McCabe, 2004).

4.4.6 Outlier removal

In particular, HR is susceptible to artefacts using the chest strap. For example, activation of the chest muscles can result in a noisy ECG trace with enough contraction. As such, any instances in time where HR is above 190, linear interpolation is used in Microsoft Excel between the previous and the next values. A max HR of 190 was chosen as this is the average maximum heart rate for the youngest age group (25-34) in our sample as depicted by the common equation: 220-age. An example of linear interpolation is shown in Table 4.4. This method uses far less computational power than using MATLAB, where more code is required to process the same command, using the 'MATLAB live script' function. Removal of the outlier prevents drastic changes in the chromatic transformation of the signal, which is explained in detail in section 2.4.4. If there are a significant number of outliers in a short duration, this will be removed from analysis.

Table 4.4: Time variation with an outlier of 204 at 08:02:00

Time (24 hour)	Example with outlier	With linear interpolation
08:01:00	102	102
08:02:00	204	117
08:03:00	132	132

If the HR parameter is linearly interpolated, so is the HRV measure (RMSSD), as the RMSSD is dependent upon the HR values. No outliers are removed from the other parameters in this methodology.

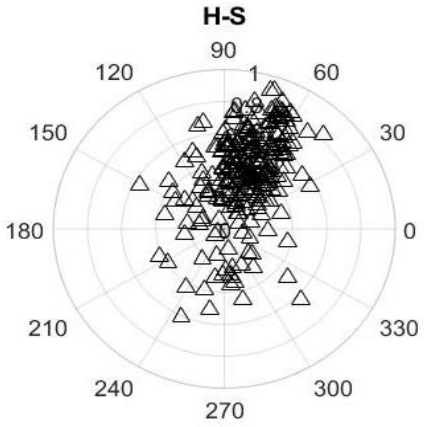
4.5.1 Chromatic filtering for proposed advisory system

The chromatic analysis is performed in Microsoft Excel as the spreadsheet format allows clear visual representation of large data sets. As explained in section 2.5, chromatic analysis includes three or more filters, which can be of varying widths depending on the application. Similarly, the frequency of time stepping may be chosen. The way these variables may be manipulated for optimal algorithm strength are explained in this section.

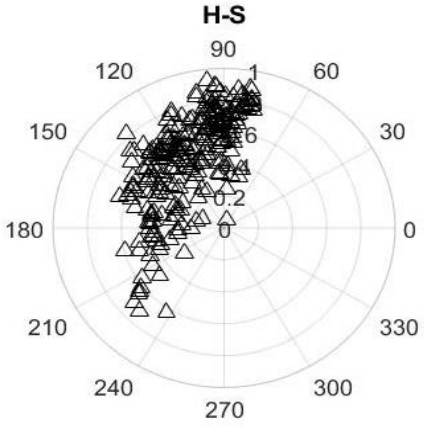
4.5.2 Filter choice for PSD

In order to choose the appropriate filter for the analysis of PSD (section 4.4.4.1.2), the H-S polar plots for three filter types were analysed (Figure 4.16) for one 24hr shift. The H-L plots were removed from visual analysis, due to the extreme spread observed between each discrete, 5-minute sample of PSD data (as outlined in section 4.4.4.1.2). The PSD varied significantly between samples, which may be due to noise, causing large fluctuations in the overall PSD observed from one point to another. As shown in Figure 4.16a, the triangular filters show a considerable amount of data points between 60-90° (R filter dominance), whereas Figure 4.16b shows a spread within 75-240° (G filter dominance), due to the large size of the G filter, compared to R and B filters. Figure 4.16c, using the modified truncated filters, greatly reduces the amount of G dominant filters, which is more meaningful between discrete data sets, more easily establishing period of sympathetic (R dominance) and parasympathetic (G or B dominance).

a)



b)



c)

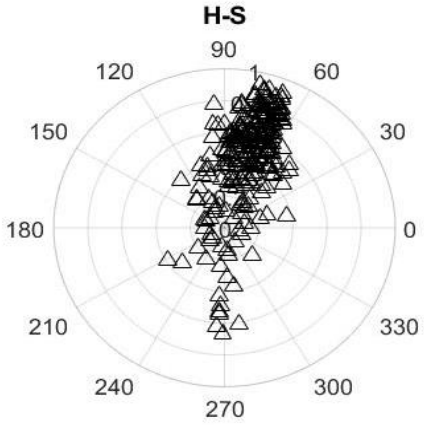


Figure 4.16: Visual interpretation of H-S polar plots for a) triangle filters, b) truncated filters c) modified, truncated filters with short green filter half the height of b.

4.5.3 Filter width

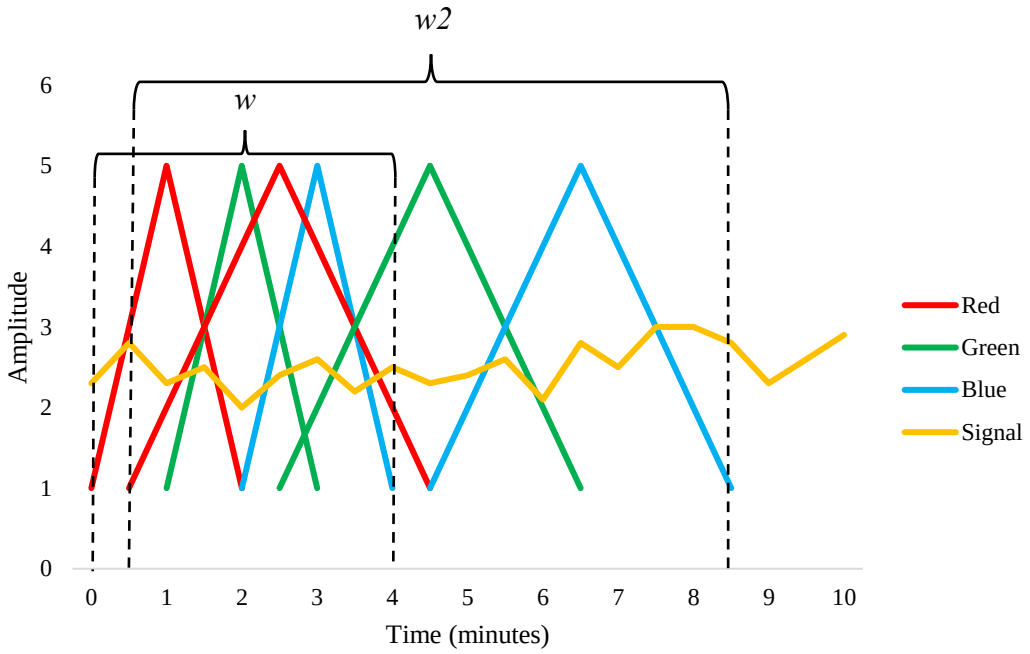


Figure 4.17: Examples of triangular filter widths on an amplitude : Time diagram

The width determines the length of time per filter and the spread over time of all three filters collectively. As shown in Figure 4.17, the total width, *w*, spans four minutes, *w2* spans eight minutes. Both types of filters are the same amplitude, however one set of filters is twice the width per filter and overall. The filter width will be analysed at different lengths to identify the best predictor of fatigue over time.

4.5.4 Number of filters

Most applications of chromatic filtering use three filters, *R*, *G* and *B*. Figure 4.18 shows five triangular filters, where filters *D* and *E* are between the previously described three-filter approach. The additional filters can lead to more detailed analysis but at additional computational cost. The number of filters will be examined for optimal analysis whilst maintaining minimal computational cost.

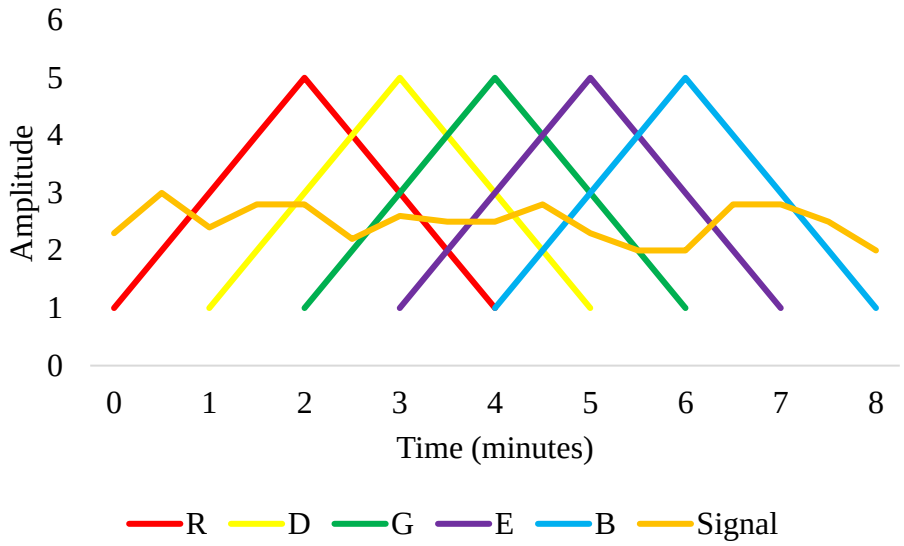


Figure 4.18: Example of 5 chromatic filters R, D, G, E, B versus time

Where four filters are analysed, two sets of chromatic transformation (section 2.4.3) are applied. *H, L* and *S* parameters are analysed from filters *R, G* and *B*, as well as filters *R, D* and *B*. The *H, L* and *S* parameters from both sets of chromatic transformation are averaged, which causes a shift in the chromatic data representation on the polar plots.

Where five filters are analysed, there are three sets of chromatic transformation. Thus, *R, G, B,* and *R, D, B* and *R, E, B* are chromatically transformed. The results from *H, L* and *S* are averaged between the three sets of data, resulting further shift in chromatic data representation. By using more than 3 filters, signal information quantification can be enhanced with the use of additional filters (Jones et al., 2005).

4.5.5 Parameter normalisation

Each variable is set parameters for normalisation (see section 2.4.3) which can be modified either to the individual or to the shift type. Using the HR estimator (Table 4.5), each participant can have a different maximum parameter depending on their age group, where the average age per group is used:

Table 4.5: HR estimator for different age groups, affecting maximum parameter

Age group	Maximum HR
25-34	190
35-44	180
45-54	170

For an individual in the age group 25-34, the 'max' parameter can be set to 190 and the 'min' parameter as 60 (Aeschbacher et al., 2017). Generally, HR does not fall outside of this range and they will be classified as outliers and linearly interpolated (see section 3.4.6). Therefore, where HR may be 104, Eqn 8 may be utilised:

$$0.338 = \frac{(104 - 60)}{(190 - 60)}$$

Changing the min/max parameters ultimately affects the strength of the signal, *L* (section 2.4.3). The parameters may be modified so that the optimum algorithm is produced to predict the KSS index on shift and effectively predict the onset of fatigue. Similarly, for parameter f. 'time on shift', the max parameter may be interchanged to the maximum length each shift worker is expected to work on their shift.

4.5.6 Chromatic analysis

The average *H, L* and *S* parameters for each participant were saved into .txt files and loaded into MATLAB. MATLAB was used to produce the polar plots where necessary, for data representation.

4.6 Randomisation of participant groups

Using a random number spinner, the first group of participants (6) was selected. Table 4.6 shows the two groups, as well as the relevant shift pattern and gender. Group 1 shows the order as chosen by the random number spinner, group 2 shows the residual participants who were not chosen. As demonstrated, an even gender split was obtained for the participants.

Table 4.6 Two groups of participants, where group 1 has the data fitted and group 2 acts as a validator of the algorithm.

Participant number	Gender	Shift type
Group 1		
15	Male	Night
2	Female	Night
6	Male	Day
3	Male	24 hour
14	Female	24 hour
12	Female	Night
Group 2		
1	Male	24 hour
8	Female	24 hour
9	Female	Day
10	Male	Day
11	Female	Day
13	Male	Night

4.7 Matrix system for the identification of rest without user input

As explained in section 4.3.1, component ‘f’ looks at the rise of the time on shift variable, which is manipulated by sleep quality and the duration of rest. However, the aim of the algorithm is to minimise the user input as much as possible. Firstly, what ‘rest’ consists of is to be debated. Nejati et al. (2016) consider that restorative break programmes do not necessarily mean cessation of physical activity, but rather an opportunity to relax by taking on a different mind-set.

A meta-survey on the nature of meaningful, restorative rest concluded that rest emerges from the engagement in activity that is personal, quiet and effortless (Nurit and Michal, 2003). When asking the participants to log their breaks, it is not known what activities they are performing during this period without watching the participant. However, it may be possible to identify periods of restoration, low or high fatigue by producing a matrix looking at the association between (a) vector magnitude (VM) from accelerometer data and using (b) Power Spectral Density (PSD) of the frequency-

domain heart rate variability (HRV) measures. Therefore, in accordance with the 5-minute PSD measurements, the average VM every 5-minutes and PSD produce a matrix, which is explained by the variables in Table 4.7 and 4.8 below.

(a) VM

Whilst wearing the equipment before testing on the participants, three classification types for the activity of an individual during their shift were identified, as shown in Table 4.7 below:

Table 4.7: The classification of three stages of activity to be computed into a matrix of fatigue onset.

VM threshold	Activity
0-10	'Stationary' – the participant is completely still, either performing surgery or sleeping, with minimal chest movement, accounting for the change in band circumference when breathing.
10-90	'Rest' – the participant is awake and moving, however the overall movement is limited. The participant is most likely operating at a desk, or stood up, but does not engage in large amounts of activity.
>90	'Active' – the participant's activity is high, at the least walking around for the majority of the 5-minutes, presumably away from their desks.

(b) PSD

By observing the dominant filters (R, G, B) of the PSD (Figure 4.17), three further thresholds are observed in Table 4.8:

Table 4.8: The classification of three stages of stress based on PSD

Hue (°) of PSD	Filter dominance
0-120 (P_3)	Red – sympathetic dominance, indicative of stress, anxiety or other negative emotional states.
121-195 (P_2)	Green – sympathetic/parasympathetic dominance, a well-balanced autonomic nervous system, able to return to a rested state quickly after stressors have passed.
196-360 (P_1)	Blue – parasympathetic dominance, associated with relaxation, calmness and positive emotional states.

Applying both Tables 4.7 and 4.8, produce a matrix shown in Figure 4.19.

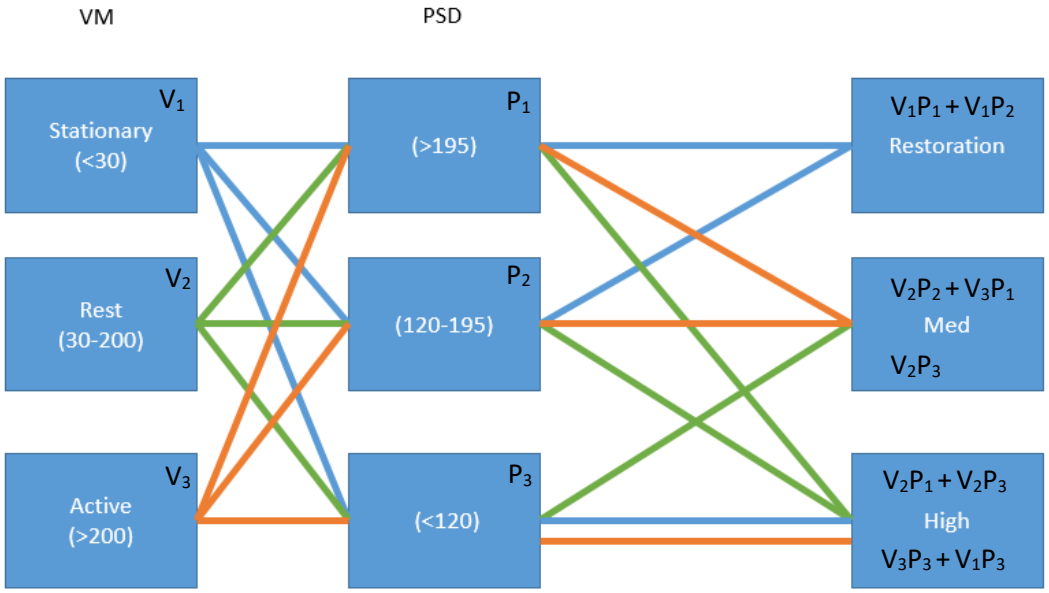


Figure 4.19: Matrix of fatigue index as outlined by VM and PSD (The angle of Hue) parameters to show the restorative or accumulation of fatigue.

Figure 4.19 shows scenario's in the workplace, which are indicative of restoration (stationary VM and G or B dominant filter), and medium or high fatigue accumulation. The restoration event is a true example of when a participant is in a relaxed state with little stressful stimuli for at least 5-minutes, providing beneficial and restorative effects, this is apparent for both G or B dominant filters, where there is more parasympathetic dominance occurring. The high fatigue accumulations are in situations where the

participant is stationary with R dominant filters, or when resting with B dominance. These scenarios may be indicative of surgery, where the participant is stationary with sympathetic dominance, indicative of stress, or when the participant is performing monotonous work, i.e. sat at their desk, both of which accelerate the accumulation of fatigue (G or B dominant filters). Similarly, if the participant is engaging in a lot of physical activity and their PSD indicated R dominant filter, this is a high stressor scenario in terms of the physical accumulation of fatigue. The other states are indicative of a medium accumulation of fatigue. All scenarios where activity is high is medium risk, as the participant is engaging in physical activity and therefore less susceptible to the short-term risks of fatigue often seen with monotonous work, as outlined by the pilot study (section 3.3) and aforementioned by Hockey (2011) in section 2.2.1. Being in a restful state with R dominant filters are outlined as medium risk as the participant is engaging in activity which is providing stimulus to the brain or muscles, reducing the perceived monotony of the task at hand.

In accordance with the formula outlined in Table 4.1, medium risk follows the rise outlined in this table. For categories of high risk, the rise is double that of the medium risk, for the five minute period. If the rise were to increase by 0.02 per minute in hour 16 on shift, it would instead increase by 0.04 per minute. Research conducted by Horne and Reyner (1996) and Reyner and Horne (1997), found that caffeine and a <15-minute nap effectively reduce sleepiness in drivers by 1hr. This relationship of 1:4 (1 minute of rest = 4 minutes of wakefulness) is used for the restoration event rate in the matrix. Whilst the ingestion of caffeine before breaks was not instated in this line of work and the participants are not driving, there is currently limited evidence as to the effects of breaks and the subsequent restorative effects on an individual in the workplace. Future research will need to look at these parameters and matrices to further improve the algorithm. For the purposes of this study, each 5-minutes of rest outlined by the restoration in Figure 4.19, 20-minutes is reduced from the time on shift variable, relative to the hour of which the rest is taken. For example, if 5-minutes of rest is logged at hour 4, then 5 samples of 1 minutes' worth of data (as per Table 4.1) is multiplied by 4 (the 1:4 ratio), then divided by 5 to give the reduction in the time on shift variable each minute. E.g. ((0.005*5)*4)/5 = 0.02. Therefore, each 5-minutes of rest observed by the matrix, results in a reduction of the time on shift variable of 20 minutes' worth

of medium-risk fatigue accumulation. The effects of the outcomes of the matrix are outlined in Table 4.9.

Table 4.9: The increase in the time on shift component of the algorithm per minute, based upon PSD and VM, respective of the hour into the shift.

Time into Shift (Hours)	Medium	High	Rest
0	0.00125	0.0025	-0.005
4	0.005	0.01	-0.02
8	0.01	0.02	-0.04
12	0.015	0.03	-0.06
16	0.02	0.04	-0.08
20	0.025	0.05	-0.1
24	0.03	0.06	-0.12

The chromatic results chapter will look at the association between both algorithms, with and without the matrix, to see whether it serves as a good proxy for the identification of strategic rest periods/breaks on shift. If successful, it further mitigates the requirement for user input into the algorithm when stating that they are on a break, as it is not known what activity the participant may be up to, or perhaps they forgot to clock off on the break as per the questionnaire outlined in section 4.2.2.1.

4.8 Summary

This chapter covers the methodology for the second round of testing with our cohort of surgeons for the monitoring of fatigue in the workplace, using the KSS (section 2.2.1.2). The proposed model for identification of workplace fatigue and its components are explained, which include HR, HRV, ST, time on shift and predicted KSS, as well as the data analysis techniques and analytical platforms. The statistical analysis in SPSS, removal of outliers and chromatic filtering components are also described.

The justification for each variable is described in this section, as well as the equipment used. The randomisation of participants is shown, which will be used to fit the algorithm which is described in chapter 6. The matrix system is described, providing a unique alternative to logging the duration of rest on shift, mitigating the possibility of user input error in the algorithm, based on PSD and VM. All data in the next chapter includes all participant's data for the justification and verification of the variables to be included in the final algorithm.

Chapter 5

Experimental Results and analysis

5.1 Introduction

The following section includes an analysis and discussion of the collected questionnaire and chest strap data retrieved from all participants. The section will provide statistical analysis on the KSS variance across shifts and genders, as well as the correlation between the biomarkers. Following on, this section will include more statistical analysis into the components which are to be carried forward for chromatic analysis (chapter 6). The justification for exclusion of irrelevant variables, which will take up additional computational cost and worsen the effect of the algorithm are also explained.

In total, for the 12 participants (excluding three due to missing data), there are 9557 sets of data for every variable, each set of data represents 1 minute. Data is 'trimmed' so that the total collected data is comparable for every variable, namely: Karolinska Sleepiness Scale (KSS), Heart Rate (HR), Skin Temperature (ST), Heart Rate Variability (RMSSD) and Accelerometer (ACC), Time on Shift and Predicted KSS. LF/HF ratio was excluded from analysis due to being averaged every 5 minutes and not every minute as the other variables were. For example, if participants began recording KSS values before the equipment was worn, or a participant had removed equipment but carried on filling out KSS, data is only included where there are comparable data sets for all variables.

5.2 Surgeon fatigue questionnaire results

5.2.1 Introduction

This section will discuss the responses and outcomes of the questionnaire administered to the surgeons (section 4.2.2.1). Similarly, sleep quality will be addressed in this section. The questionnaire was sent every three hours on shift. In total there were 65 text messages sent for participants to fill out the questionnaire however, 59 were answered (91%).

5.2.2 Surgeon fatigue questionnaire analysis

Every participant answered multiple questionnaires, which shows the fluctuations of the KSS for each participant throughout their shift. As a result, statistical analysis observes trends in the data. Average KSS values in males and females were 4.22 $\pm$ 1.43 and 3.31 $\pm$ 1.75, respectively; the distribution in the two groups differed significantly (Mann-whitney U = 6219753, z = -31.59, p < 0.001). Kruskal-Wallis test identified that the average KSS values differed significantly between shift types $X^2(2) = 1209.483$, p < 0.001, where the day shift participant scored lower than the night shift and the 24hour shift workers. Table 5.1 shows descriptive statistics for KSS, split by shift type.

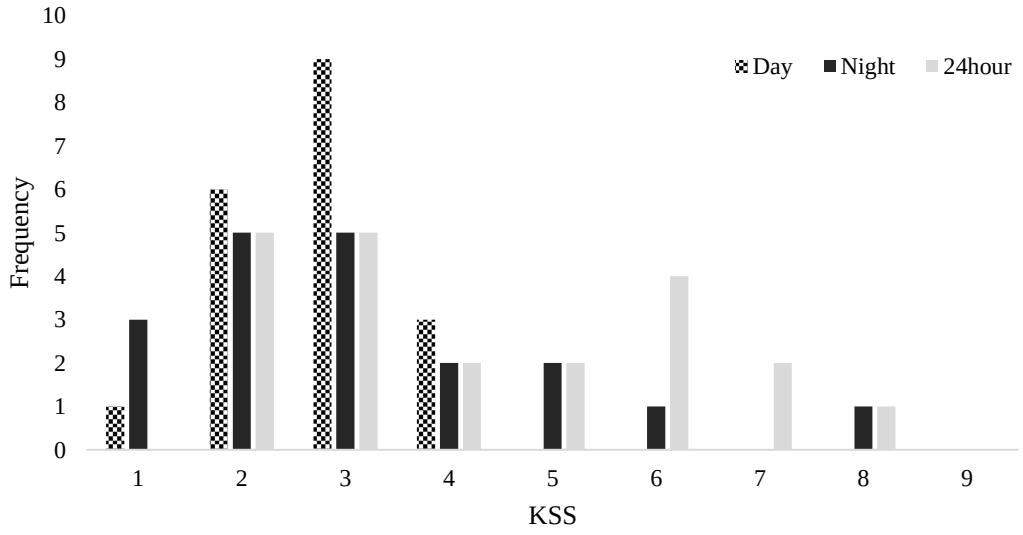


Figure 5.1: Frequency and distribution of KSS scores, split by shift type.

Table 5.1: Shift type and average KSS scores for all surgeons.

Shift type	Average	Maximum	Minimum
Day	2.75 $\pm$ 0.69	5	1
Night	3.61 $\pm$ 1.93	8	1
24 hour	4.40 $\pm$ 1.64	8	2

Figure 5.1 shows the frequency and distribution of KSS scores, split by shift type. In total, there were 59 questionnaires answered, with 59 KSS scores.

Participants on day shifts were not subject to fatigue, as their KSS score did not reach six. However, participants on night and 24 hour shifts did include examples of instances where surgeons were likely to be fatigued and potentially cause errors. Overall, there were nine recorded instances of participants being subject to fatigue.

Figure 5.2 highlights the frequency of KSS scores before a participants' break and afterwards. It can be seen that participants who reported higher KSS values benefitted from breaks more than those who reported two or three, whereby KSS scores did not change after a break. Breaks were more effective at reducing KSS when participants were already fatigued ($\geq$6 KSS), Mann-whitney U = 33, z = -2.657, p = 0.030.

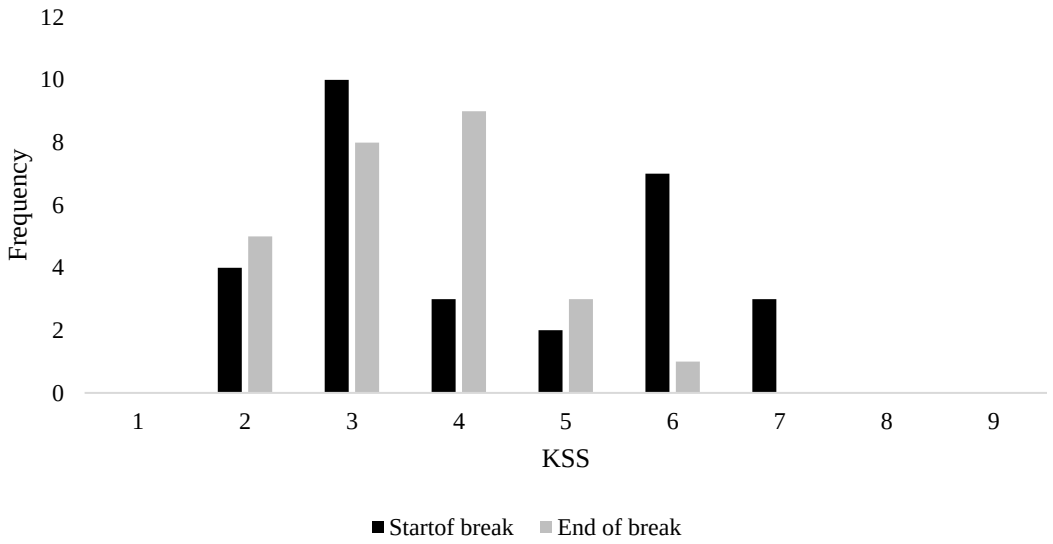


Figure 5.2: Frequency of KSS score at the start and end of breaks for all participants.

For errors made, there were 53 reports that no errors were made from the last questionnaire. However, there were one near-miss and two unsure responses (Table 5.2).

Table 5.2: Reports of near-miss or unsure errors, relative to the shift and time of day.

Type of error	KSS	Shift type	Time of day (24 hour)
Near-miss	6	Night	20:30
Unsure	7	24hour	23:00
Unsure	4	Day	17:00

No participants smoked in this experiment. Therefore, smoking is removed from analysis in this report. Further information on caffeine and naps are included in section 5.4.1.

5.2.3 Sleep quality

Table 5.3 shows the average sleep in hours for each sleep quality score, in reference to the last full sleep prior to the beginning of their shift. No participant had a shift within 24 hours of the shift being recorded. There were no significant differences in sleep quality, duration, or the KSS values shown in Table 5.3 by shift types or gender. As can be seen, the majority of participants had good sleep quality, which is coupled with an adequate amount of sleep. A quarter of our sample had average or poor sleep, which is likely to predispose participants to fatigue earlier on in their shift. This is achieved with >1 hours less sleep per night and shown by the KSS peak achieved on average for each sleep group. Half of the 'good' sleep quality was from day shift workers, explaining the reduced KSS average at its peak. KSS at the start of the shift was higher for average sleep compared to good sleep.

Table 5.3: Average sleep in hours for each sleep quality score and the relevant KSS scores at the start and at its peak during the shift.

Sleep Quality	Quantity	Sleep (hours)	KSS start	KSS peak
Good	9	7.78	2.86	4.29
Average	2	6.00	4.33	7.33
Poor	1	6.5	2	8

To provide sustenance to the hypothesised increase in y-axis shift depending on sleep quality mentioned in section 4.3.1 (component g), the difference between the start KSS and the peak KSS doubles with each stage of sleep quality (~1.5, 3, 6). As the KSS ranges on a Likert scale from 1-9, the 'good' sleep scores begin at the minimum threshold as defined by the min and max parameters. The 'average' sleep score begin at 33% of the maximum parameter as the KSS value increased by 33% of the Likert and similarly the 'poor' sleep scores begin at 66% of the maximum parameter as the KSS value increases by 66% of the total spread of the scale

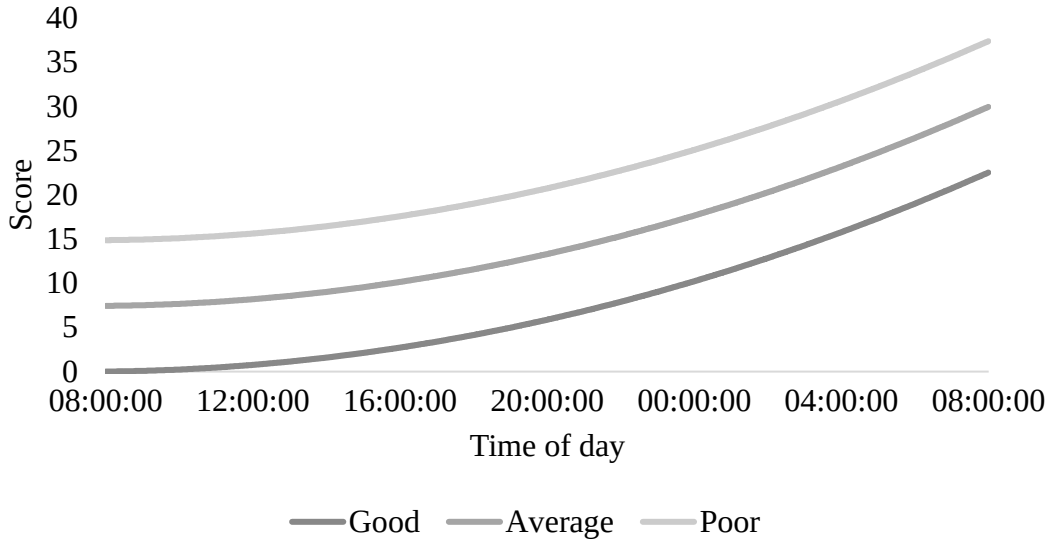


Figure 5.3: Comparison between 'good', 'average' and 'poor' sleep quality and its effect on the Time on Shift variable.

Figure 5.3 shows the difference between 'good', 'average' and 'poor' sleep quality prior to their shift. The ratings of good to poor are subjective and thus should be used cautiously. Participants who claimed to have a 'good' sleep prior to their shift, adhered to the rise in Time on Shift depicted in Figure 4.10. Participants who logged 'average' sleep had the same curve, however the value begins at 33% of the pre-determined max parameter respective of shift type, instead of 0, using the formulae: (max*0.33) + *n* (see Table 4.3 for max), where *n* is the raw value of Time on shift for the appropriate amount of minutes the participant has been on shift. For participants who had a 'poor' sleep, the score began at 66% of the max parameter, using the formulae: (max*0.66) + *n*. For example, a 24hr shift participant who had 'good' sleep quality before their shift would be at work for 1165 minutes before achieving the same score as a participant who had 'poor' sleep quality.

5.3 Chest strap results

5.3.1 Introduction

This section includes a brief analysis on the biomarkers obtained from the study and their correlations against each other. Any significant differences between shift types or genders are included in this section also.

5.3.2 Chest strap analysis

Kruskal-Wallis analysis identified significant differences for all biomarkers across the shift types. Heart rate (HR) was significantly lower for 24 hour shifts (96.26 BPM) compared to day shift (105.27 BPM) and night shift (105.27 BPM) $X^2(2) = 371.80$, $P < 0.001$. Heart rate variability (RMSSD) measures were significant higher for 24hour shifts (29.29ms) compared to day shift (26.36ms), $X2(2) = 955.56$, $p < 0.001$, but not significantly different to night shifts. Skin Temperature (ST) was significantly lower for the day shift (33.11) compared to 24hour shifts (33.90) $X2(2) = 708.54$, $p < 0.001$ and night shift participants produced higher vector magnitude values (256.00) compared to 24hour shift (193.16) $X2(2) = 162.73$, $p < 0.001$.

Table 5.4 shows the R-value (correlation) between the biomarkers included in the study. All values reached statistical significance and are compared against 9049 values, which are recorded in total. Of note, the only significant finding is the inverse relationship between HR and HRV, which is expected. That is, as HR increases, HRV decreases. The other variables have weak correlations against each other.

Table 5.4: Correlation between biomarkers (heart rate, heart rate variability, skin temperature and vector magnitude) in the study.

Biomarkers	HR	HRV	ST	ACC
HR	1	-0.549	-0.216	0.205
HRV	-0.549	1	0.140	-0.113
ST	-0.216	0.140	1	-0.110
VM	0.205	-0.113	-0.110	1

When split by gender, similar correlations are produced and displayed in Table 5.5:

Table 5.5: Correlation between biomarkers (heart rate, heart rate variability, skin temperature and vector magnitude), split by gender.

Male				
Biomarkers	HR	HRV	ST	ACC
HR	1	-0.376	-0.440	0.279
HRV	-0.376	1	0.209	-0.075
ST	-0.440	0.209	1	-0.124
VM	0.279	-0.075	-0.124	1

Female				
Biomarkers	HR	HRV	ST	ACC
HR	1	-0.569	-0.081	0.144
HRV	-0.569	1	0.143	-0.132
ST	-0.081	0.143	1	-0.102
VM	0.144	-0.132	-0.102	1

For males, the correlation between ST and HRV is stronger, whereas for females the correlation between HRV and HR is stronger. In general, most correlations are similar and either show a good correlation, or weak correlation, when split by shift type as shown in Table 5.6. The main consistency between shift type is between HR and HRV.

Table 5.6: Correlation between biomarkers (heart rate, heart rate variability, skin temperature and vector magnitude), split by shift type.

Day Shift				
Biomarkers	HR	HRV	ST	ACC
HR	1	-0.761	-0.236	0.194
HRV	-0.761	1	0.047	-0.102
ST	-0.236	0.047	1	-0.081
VM	0.194	-0.102	-0.081	1

Night Shift				
Biomarkers	HR	HRV	ST	ACC
HR	1	-0.642	-0.258	0.239
HRV	-0.642	1	0.232	-0.191
ST	-0.258	0.232	1	-0.133
VM	0.239	-0.191	-0.133	1

24-hour shift				
Biomarkers	HR	HRV	ST	ACC
HR	1	-0.524	-0.153	0.186
HRV	-0.524	1	-0.104	-0.011
ST	-0.153	-0.104	1	-0.101
VM	0.186	-0.011	-0.101	1

5.3.3 Chest strap and questionnaire discussion

As expected, night shift and 24hour shift workers reported higher KSS values than day shift workers, presumably due to working overnight and suffering from misaligned circadian rhythms. This can be justified by the frequencies of KSS responses depicted in figure 5.1. Participants had greater success at reducing their KSS values taking breaks when subject to fatigue/sleepiness. Whilst it is understood that participants take breaks as and when they can, there may be additional benefit to the organisation and individual to take breaks at alternative times on their shift.

For errors made and sleep quality, they are hard to admit. These variables may not be reported truthfully, as there is undoubtedly an unconscious bias that making errors is bad for the individual and the organisation. However, if true, our cohort have been very successful at mitigating errors despite 50% of our cohort reaching a 6 on the KSS. 24hour and night shift workers may not have good sleep quality prior to their shifts due to misaligned circadian rhythms. There is no objective data collected for sleep quality to see what time participants were sleeping and whether it is an unnatural time as expected with the day shift workers. Sleep quality can be hard to depict with a Likert scale, especially if one's perceptions of 'poor' sleep are different. Nonetheless, there is a trend that average or poor sleep results in a larger KSS peak during the shift.

HR was significantly lower for 24hour shifts, which may be due to sleep opportunities taken that therefore reduces the overall HR for significant durations on the shift. Similarly, HRV measures are higher for 24hour shifts, as HRV is inversely proportional to HR. ST is susceptible to artefacts, such as pressure, which can be applied to the temperature sensing pod that increases the temperature. As the pressure cannot be measured, ST can vary considerably between participants. Similarly, whilst all participants wore surgical scrubs, some participants were wearing additional layers underneath their scrubs which can increase the temperature further. ACC was significantly higher for night shifts compared to 24hour shifts, but similarly to HR and HRV, there are large periods of inactivity for 24hour shift workers who had opportunities for sleep on their shift. As a result, this reduces the average vector magnitude per minute.

5.4 Algorithm Statistical Analysis

5.4.1 Break algorithm

Table 5.7 shows the habits identified on surgeons breaks. Smoking is not included as no participants in the study smoked. Overall, there were 26 breaks reported for the 12 participants.

Table 5.7: Description of each break recorded from all participants, in ascending order for duration of break.

Duration of break (minutes)	Decrease in KSS	Nap	Caffeine	Shift type	Gender
10	0	No	No	Day	Male
10	0	No	Yes	24hrs	Male
15	0	No	Yes	24hrs	Male
15	0	No	Yes	24hrs	Female
15	0	No	Yes	24hrs	Female
15	0	No	Yes	Day	Female
15	0	No	Yes	Day	Female
15	0	No	No	Day	Female
20	0	No	Yes	24hrs	Male
20	0	No	Yes	Night	Female
20	0	No	No	Night	Male
25	1	No	No	Day	Male
25	0	No	No	Day	Female
30	0	No	No	24hrs	Male
30	2	No	No	24hrs	Male
30	1	No	No	Night	Female
30	0	No	No	Day	Male
30	0	No	No	Day	Male
30	0	No	No	24hrs	Male
30	0	No	No	Day	Male
30	1	No	No	Night	Male
60	1	No	No	Night	Male
90	0	No	No	24hrs	Female
150	3	Yes	No	Night	Male
350	2	Yes	No	Night	Female
370	4	Yes	No	24hrs	Male

One-way AVOVA identified no significant differences in caffeine occurrences by shift type. Similarly independent samples t-test identified no significant differences in caffeine or naps by gender, which provides homogeneity across our sample of participants.

Figure 5.4 shows the relationship between time on break and the reduction in KSS value from pre to post break.

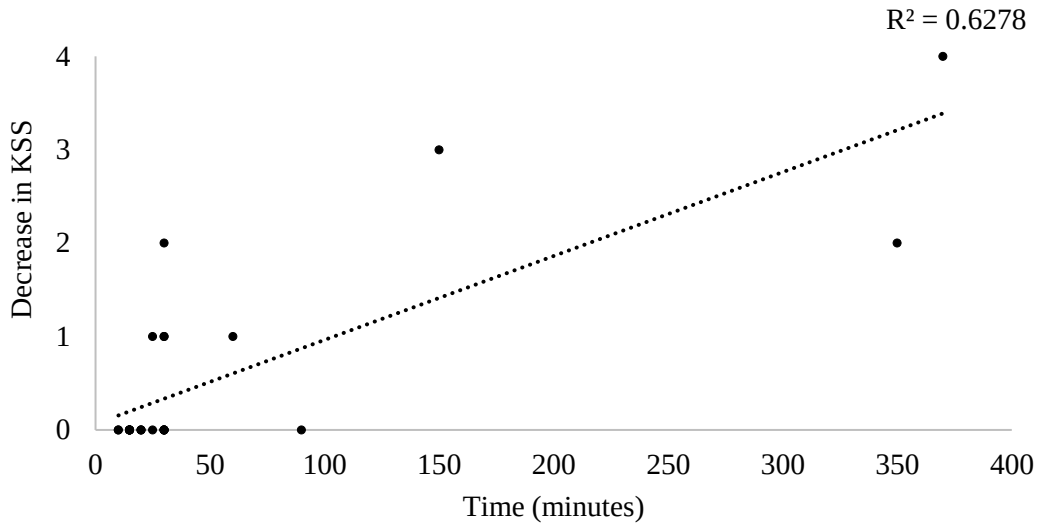


Figure 5.4: Relationship between length of break and reduction in KSS for all participants, as depicted in Table 5.7.

An increased break length significantly correlates to a decrease in KSS, r(26) = 0.792, p <0.001. Break length was regressed on predicting the decrease in KSS. Moreover, the $r^2 = 0.628$ depicts that the model explains 62.8% of the variance in KSS.

The line of best fit depicted in figure 5.3 has the gradient:

$$y = 0.07 + 0.00898*X$$

with reference to the *x* and *y* axis in figure 5.4.

5.4.2 KSS correlation

The correlation between KSS and the four biomarkers are displayed in the scatter graphs below (Figures 5.5 - 5.8). The r^2 values and lines of best fit are also shown in the plots.

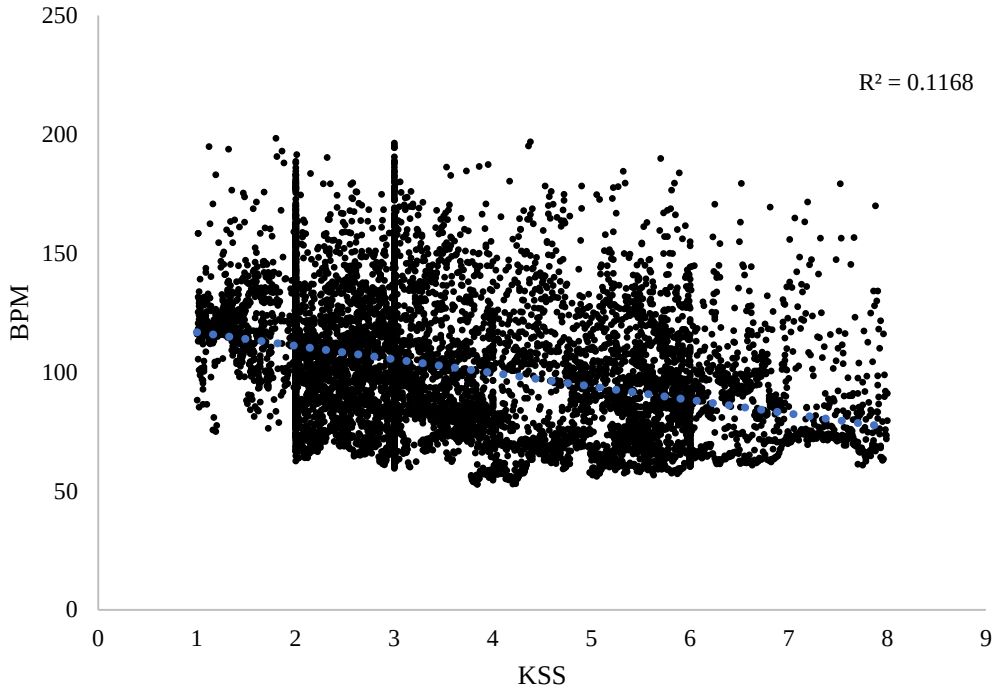


Figure 5.5: Scatterplot of correlation between HR and KSS for all participants.

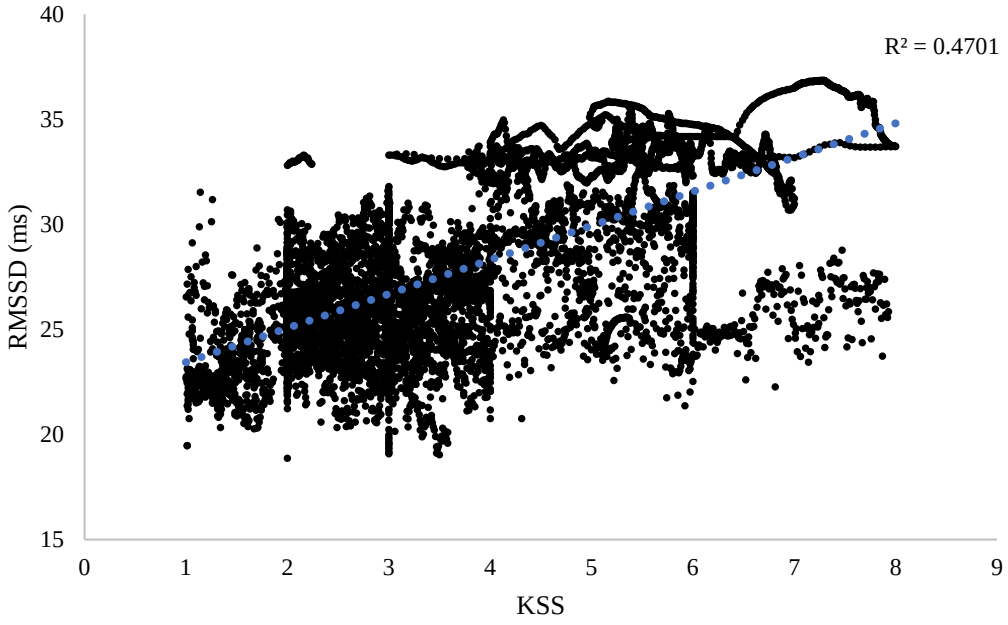


Figure 5.6: Scatterplot of correlation between RMSSD and KSS for all participants.

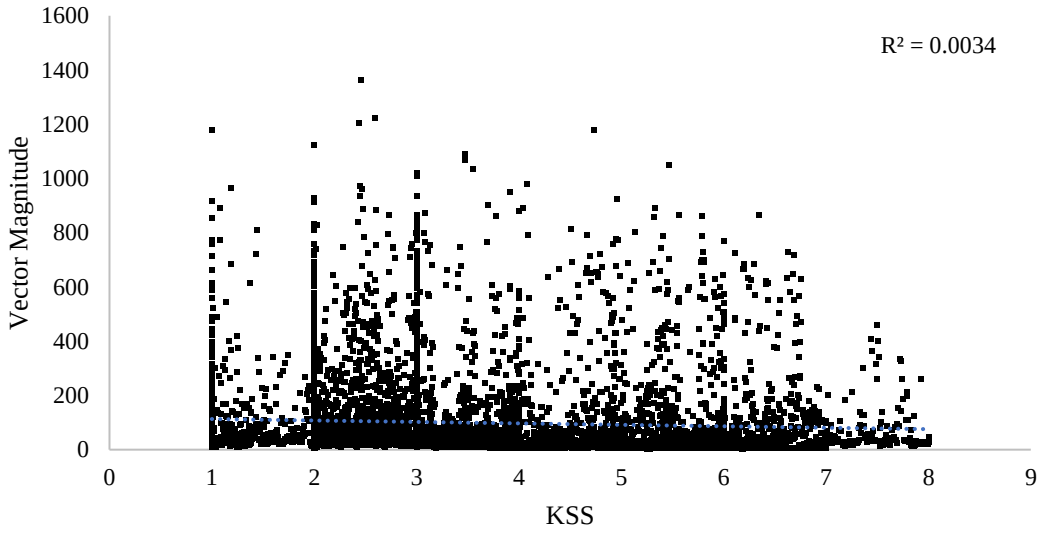


Figure 5.7: Scatterplot of correlation between Vector Magnitude (VM) and KSS for all participants.

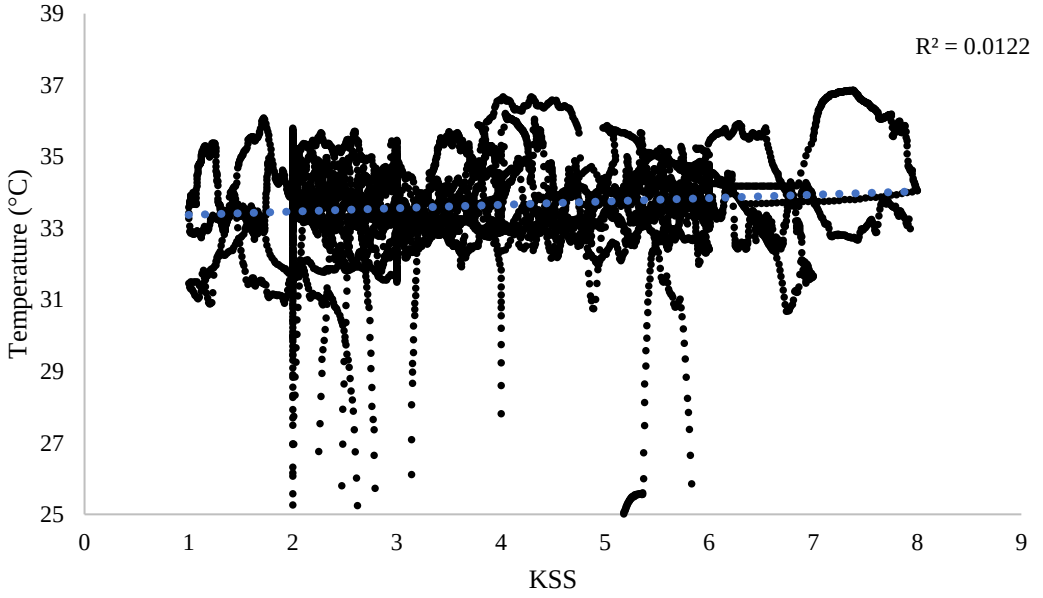


Figure 5.8: Scatterplot of correlation between Skin Temperature (ST) and KSS for all participants.

Heart rate and KSS had a moderate negative correlation, $r(9557) = -0.332$, $p < 0.001$. RMSSD had a near strong positive correlation, $r(9557) = 0.697$, $p < 0.001$, whereas VM and ST data had weak correlation ($r(9557) = -0.058$, $p < 0.001$, and $r(9557) = 0.090$, $p < 0.001$, respectively). As can be shown, 47.01% of the variation of KSS can be explained by RMSSD. The cumulative sum of RMSSD and VM were also analysed against the KSS scores, $r(9557) = 0.599$, $p < 0.001$

and r(9557) = 0.445, p < 0.001, respectively. These both show a moderate correlation, however, the regression analysis identified only 35.8% and 19.8% of the variance in KSS.

Table 5.8 and 5.9 show the differences in r^2 values when split by gender and shift type:

Table 5.8: r^2 values for biomarkers (heart rate, heart rate variability, skin temperature and vector magnitude), split by gender

Biomarker	Male	Female
HR	0.068	0.079
HRV	0.473	0.436
ST	0.013	0.010
VM	-0.042	-0.041

Table 5.9: r^2 values for biomarkers (heart rate, heart rate variability, skin temperature and vector magnitude), split by shift type

Biomarker	Day	Night	24hour
HR	0.042	0.281	0.070
HRV	0.005	0.740	0.263
ST	0.047	0.014	0.043
VM	-0.061	-0.016	-0.041

When split by gender, there are no dissimilar results between the biomarkers association with KSS. However, when split by shift type, there is a clear difference in relationship for the HRV measure across the shifts. Night shift participants account 74% of the KSS variance to HRV, in comparison with Day shift workers, that can only account 0.5% of the KSS variance to HRV.

5.4.3 Algorithm discussion

With reference to Figures 5.5 - 5.8, it is concluded that ST and VM have a weak association with KSS variance. In this study the link between ST, VM and KSS

cannot be inferred, therefore they are removed from any further analysis in this study with regards to Eqn 9 (see below). For variance of VM, it is expected that all participants are constantly moving as part of their job, which is why vector magnitude is fairly consistent across the KSS. ST is fairly uniform and consistent, as participants are unlikely to face large variations in temperature. It is also important to note that ST changes are susceptible to artefacts when there is external pressure applied to the sensor, for example when leaning on the sensor which can increase the temperature of the pod.

As mentioned before, KSS values $\geq$6 are associated with signs of sleepiness. Similarly, section 5.2.2 identified that breaks are most effective when participants score $\geq$6 on the KSS. Therefore, an alarm/break recommendation should aim to predict when a participant is at a six on the KSS scale, to achieve a greater reduction on the KSS and prevent the participant from scoring a six or above.

HR and HRV are two measures which explain over half of the variance in the KSS. However, HRV is calculated from HR, yet yields a much more significant relationship for KSS. To reduce the computational cost of the proposed algorithm, HR will be removed from further analysis.

Overall, the proposed algorithm to predict the value of KSS, will include the following equation:

$$FI = \Sigma(HRV(t) + KSS\_Pred(t) + (ToS_{SQ}(t) - DR(t))) \text{ ----------Eqn 9}$$

Where,

- FI = Fatigue index
- HRV(t) = Heart Rate Variability

- KSS_Pred(t) = Predicted KSS with respect to time
- ToS_{SQ} (t) = Time on Shift based on sleep quality with respect to time
- DR(t) = Duration of rest with respect to time

In the next chapter, Eqn 9 will be compared again Eqn 10 (below), which includes modification from the matrix (using chromatic parameter H from the power spectral density (PSD) analysis and the VM) outlined in section 4.7.

$$FI = \Sigma(HRV(t) + KSS\_Pred(t) + (ToS_{SQ}(t) + MATRIX)) \text{ ----------Eqn 10}$$

Where,

- FI = Fatigue Index
- HRV(t) = Heart Rate Variability
- KSS_Pred(t) = Predicted KSS with respect to time
- ToS_{SQ} (t) = Time on Shift based on sleep quality with respect to time
- MATRIX = further increase or decrease based on criteria from section 4.7.

5.5 Summary

This section includes the results obtained from the surgeon questionnaire as well as the data from the chest straps for every participant. The section highlights the increased KSS values from night and 24hour shifts, as well as the effectiveness of breaks and when is best to take them. The sleep quality is assessed, finding those who scored average or poor had, on average, higher peak KSS values during their shifts.

The biomarkers were correlated against KSS, whereby RMSSD had a strong correlation and accounted for 47.01% of the variance in KSS. HR has a moderate negative correlation whereas ST and VM had a weak correlation. Due to the environmental elements and the activity during the shifts for all surgeons, ST and

VM are to be removed from further analysis. Similarly, to reduce computational cost of the proposed algorithm, HR will be removed from further analysis.

Chapter 6

Chromatic analysis and discussion of results

6.1 Introduction

The analysis of the results in chapter 6 include the application of the statistical results in chapter 5, using the chromatics from chapter 4 (Eqn 9 and Eqn 10, respectively). The application of chromatics assists in reducing the complexity of unprocessed, raw signals produced from the Time on Shift, Heart Rate Variability (HRV) measures and circadian rhythm variables. The information derived from the chromatic transformation can be further analysed to predict the onset of fatigue in the workplace.

The onset of fatigue can be flagged using an advisory system, which monitors the total average strength (L) of the signals based on the individual weightings of each parameter (section 5.4.2) to classify tiredness information efficiently. Polar plots and line graphs can then be used to visualise the tiredness of the individual. The three triangular and overlapping chromatic filters identify fatigue trends over a participant's entire shift. The fatigue trends are individualised by the HRV measure, shift-specific and time-specific using the predicted KSS variable. Details about the applications of the H, L and S parameters are discussed.

Taking breaks causes a slight reduction in the Karolinska Sleepiness Scale (KSS). Its application is discussed in this chapter, as well as examples of individual participants and how the advisory system correctly identifies all cases of fatigue identified by the

participants in this research. This chapter ends with a recommendation for multiple alarms per shift.

6.2 Application of L for the prediction of KSS

Upon further analysis, an algorithm can be derived from the *L* parameter and its relationship with the Karolinska Sleepiness Scale (KSS). The algorithm chosen in this chapter considered filter length, height, width, time stepping frequencies, alternative min/max parameters (section 4.4.4.4), individualised parameters depending on shift type (e.g. shift differences in HRV as shown in section 5.3.2) and the effects of breaks on shift at reducing the value of L. The final normalisation parameters and subsequent weightings deliver the highest correlation between the *L* parameter and KSS.

Table 6.1: Amendments to algorithm and subsequent outcomes between KSS and *L* parameter.

Amendments	**Outcomes**
Adjustment of min/max parameters for normalisation	Reduction in R-value (correlation)
Time on Shift linear rise	Reduction in R-value (correlation)
Modified HRV min/max parameters by shift	Reduction in R-value (correlation)
Increased height of filters	Reduction in R-value (correlation)
Increased number of filters	Extra computational cost
Increased width of filters	Smooths *L* data but delays the time for the alarm to be triggered
Sleep resets Time on Shift parameter	Increase in R-value (correlation)
Moving average of RMSSD data	Increase in R-value (correlation)

Table 6.1 shows the amendments to improve the correlation between the *L* parameter and KSS. Modifications to the number, width and height of the chromatic filters *R*, *G* and *B* resulted in a reduction of the correlation between *L* and KSS. Whilst increasing the filters' width smooths out the *L* parameter, it may delay the onset of the advisory alarm, and therefore the width chosen is nine minutes for all parameters. Adding extra filters, up to five in total, increases the computational cost of the algorithm with no benefit to the correlation between *L* and KSS. Increasing the height of the filters

significantly reduces the correlation due to the large fluctuations in HRV every minute. The HRV parameters were modified by shift type (Table 5.6). However, the R-value was significantly reduced. The circadian rhythm differentiates the shifts and even though no correlation was identified between HRV and day shifts, the risks of the alarm sounding on day shifts are lower than on 24-hour or night shifts (Table 5.1).

Finally, due to the large variability in the RMSSD measure, the data is smoothed with a moving average of 30 minutes (Figure 6.1). Over time, the average 'moves', so that each data point of the series is sequentially included in the averaging, whilst the oldest data point in the span of the average is removed.

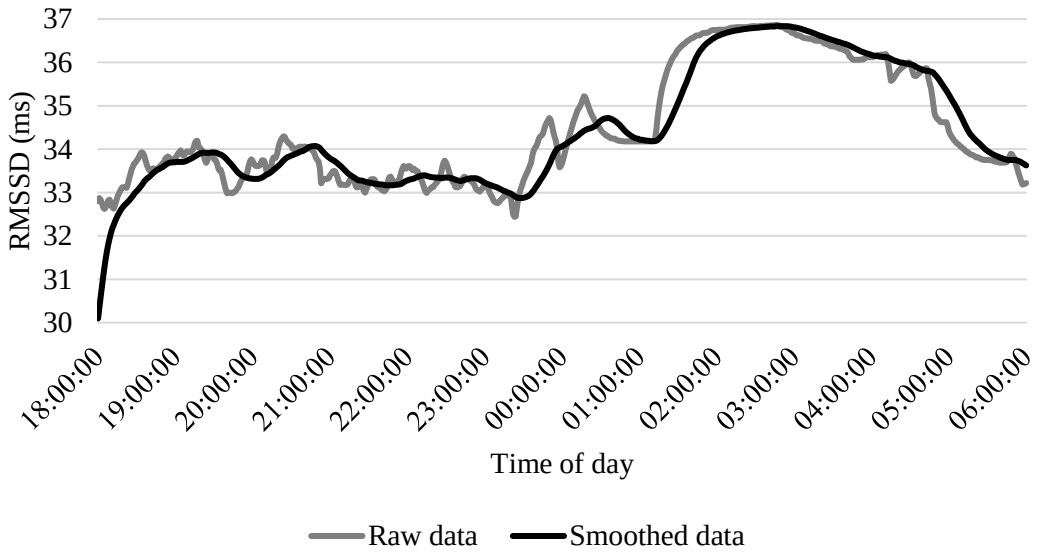


Figure 6.1: Raw RMSSD values compared against smoothed data with 30-minute moving average for one participant during a night shift.

Figure 6.2 displays the final result of the algorithm with break duration, with reference to KSS answers (Eqn 9). There is a strong correlation between KSS and L, r(5008) = 0.763, p < 0.001. The regression model explains 58.2% of the variation in KSS.

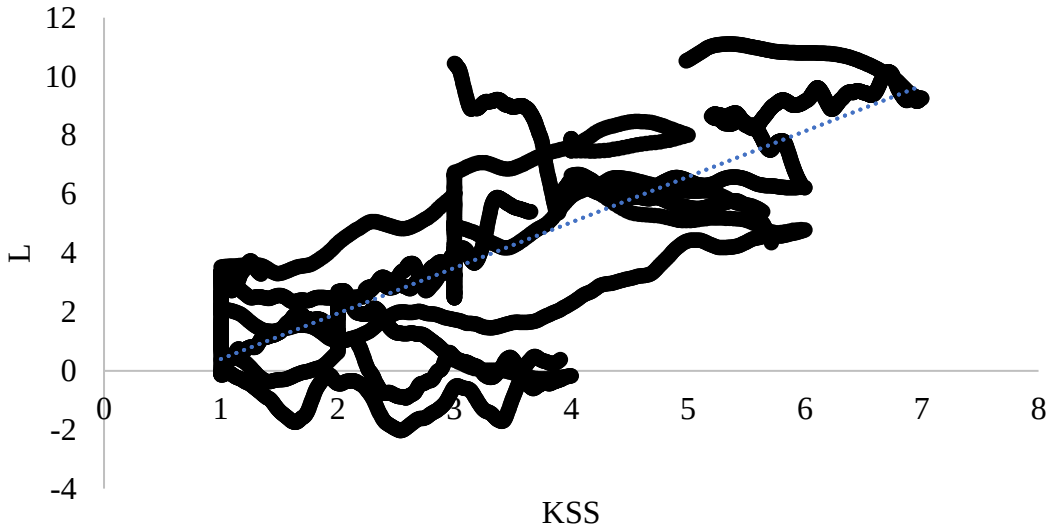


Figure 6.2: Relationship between KSS and *L* parameter for algorithm using break deductions for group 1 (Table 4.6).

The gradient of the line of best fit can be defined as:

$$y = 1.5487 * X - 1.1496$$

with reference to the *x* and *y* axis in figure 6.2.

Figure 6.3 displays the final result of the algorithm, utilising the MATRIX, with reference to KSS answers (Eqn 10). There is a strong correlation, r(5008) = 0.796, p < 0.001. The regression model explains 63.8% of the variation in KSS.

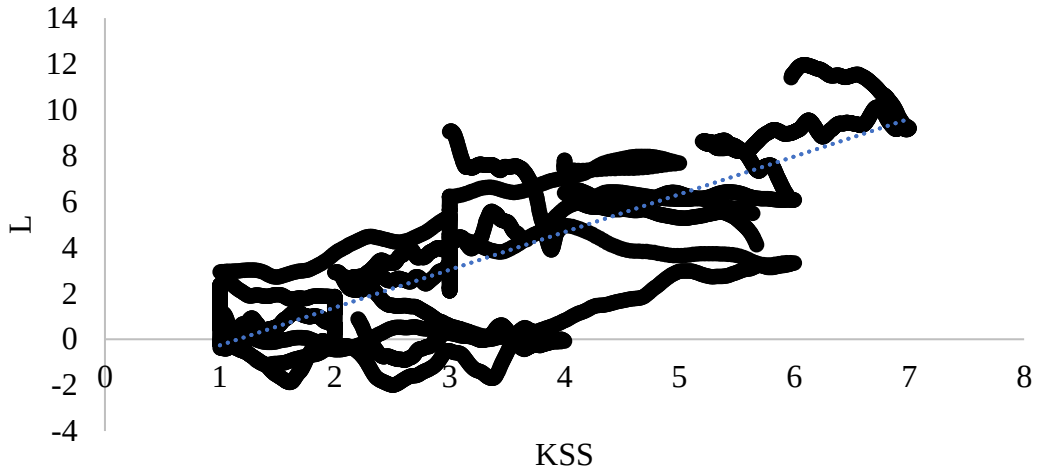


Figure 6.3: Relationship between KSS and *L* parameter for algorithm using Matrix for group 1 (Table 4.6).

The gradient of the line of best fit can be defined as:

$$y = 1.629*X - 1.827$$

with reference to the *x* and *y* axis in figure 6.3.

Table 6.2 shows the prediction of the KSS based on the value of *L*. According to the table, the alarm will trigger when the *L* value reaches an approximate value of 8.14 using Eqn 9 and 8.15 using Eqn 10 (KSS ~ 6, "some signs of sleepiness"), prompting the surgeon to take a break.

Table 6.2: Prediction of KSS from *L* parameter using Eqn 9 and Eqn 10 for group 1.

KSS	L (Eqn 9)	L (Eqn 10)
1	0.3991	-0.1641
2	1.9478	1.4988
3	3.4965	3.1617
4	5.0452	4.8246
5	6.5939	6.4875
6	8.1426	8.1504
7	9.6913	9.8133
8	11.24	11.4762
9	12.7887	13.1391

6.3 Break algorithm (Eqn 9)

Overall, 26 breaks were recorded from 12 participants. The breaks, as well as their subsequent reduction in KSS are displayed in Table 5.7. The model in section 5.4.1 shows that the break duration can explain 63% of the KSS reduction. As such, the time on break causes a linear decrease in predicted KSS using the Time on Shift variable. Table 6.3 below shows examples of the Time on Shift variable reduction based on the break length calculated in section 5.4.1.

Table 6.3: Reduction of Time in Shift variable relative to length of break taken for Eqn 9.

Break duration (minutes)	Reduction in Time on Shift variable
10	0.16
20	0.25
30	0.34
60	0.61
90	0.88

The Time on Shift variable was chosen for manipulation when breaks are administered due to its rise over time. As the breaks reduce the KSS, the reduction in the Time on Shift variable delays the onset of the advisory alarm to be triggered. In this equation, the Time on Shift variable mirrors the decline in KSS value observed due to break administration. This is to improve consistency in the algorithm across all participants.

It was observed for those shifts that had sleep opportunity on their shift that the Time on Shift variable resetting to zero upon awakening resulted in a more robust overall correlation between KSS and *L*. This is due to the significant reduction in KSS scores post-sleep compared to pre-sleep. Whilst the study did not observe the effects of post-sleep work fatigue due to surgeons waking up at the end of their shift, it is expected that if the surgeons did perform more work, they would be in a well-rested state as per KSS values. More research is required on the effects of continued work after sleeping.

6.3.2 Duration of breaks for optimal rest (Eqn 9)

Fatigue increases with respect to time (Johnston et al., 2018). A linear increase resulted in a decreased correlation between *L* and KSS (Table 6.1, Eqn 9). Figure 6.4 shows two examples of breaks, broken up into three 20-minute breaks and one 60-minute break. Due to the linear decline in breaks, it is more effective for participants to take multiple shorter breaks on shift compared to a lengthier break. As shown in Table 6.3 above, two 10-minute breaks reduce the Time into Shift score by 0.32, compared to 20-minute breaks, which reduce the Time into Shift score by 0.25. Figure 6.4 below shows that after a 12-hour shift, it takes 11 minutes more to achieve a score of >4.76 for the 1-hour-long break compared to three 20-minute breaks due to the timing of when the

breaks were taken. As a result, this will delay the time taken for an alarm to be sounded with respect to Predicted KSS and RMSSD. Further research is required to see whether the length of breaks over a shift have a significant impact on the KSS.

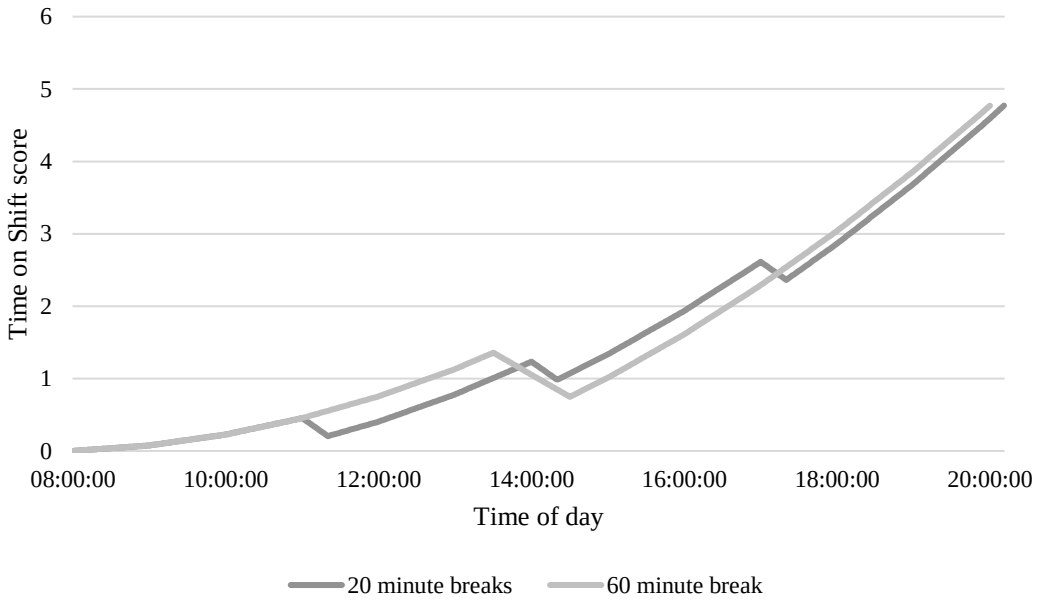


Figure 6.4: Comparison of break duration and its effects on Time on Shift score for three twenty minute breaks and one sixty minute break.

6.4.1 Interpretation of L (Eqn 9)

As mentioned in section 6.3, the administration of breaks reduces the Time on Shift variable. To advance this technique, the time on break further delays the onset of the alarm triggering by increasing the predicted *L* value relative to the length of the break, so that the alarm does not sound too frequently when participants are on shift.

As an example, a participant who records a 30-minute break before the alarm is triggered (KSS = 6), increases the alarm threshold by the gradient for the length of break and reduction in KSS (Figure 5.3):

$$0.07 + (0.00898*30) = \underline{0.34}$$

The new alarm threshold for KSS of 5 (Table 6.2, Eqn 9) would be:

$$8.14 + 0.34 = \underline{8.48}$$

If the participant takes more breaks before the alarm is triggered at the new threshold, the threshold increases further thus reducing the likelihood of alarms triggering in quick succession, should there be large variation in RMSSD measures. All subsequent alarm thresholds are increased relative to the amount of time the participant has spent on a break. If the participant does not take any breaks, the alarm threshold stays as the values in Table 6.2.

Table 6.4: Each participant's shift type included in the study.

Participant	Shift type
1	24-hour
2	Night
3	24-hour
6	Day
8	24-hour
9	Day
10	Day
11	Day
12	Night
13	Night
14	24-hour
15	Night

With the information from Table 6.4, Table 6.5 shows KSS at the point where the alarm (KSS score of six) would trigger for the first group of 6 participants, using Eqn 9:

Table 6.5: KSS score when each participant's *L* parameter would trigger the alarm (considering total elapsed time on break) using Eqn 9.

Participant	L	KSS	Time into shift (minutes)	Total elapsed time on break (minutes)	Percentage of shift before alarm
2	N/A	N/A	N/A	0	N/A
3	8.47	5.44	546	30	43.58
6	8.40	4.55	547	10	114.20
12	N/A	N/A	N/A	30	N/A
14	N/A	N/A	N/A	150	N/A
15	8.91	5.08	636	75	75.09

As shown in Table 6.5, the alarm would have sounded for 3/6 participants. What is important is reducing the amount of false alarms for participants who are not fatigued. All day shift participants did not score KSS $\geq$6, therefore the alarm should not sound for these participants. However, it triggers for participant 6, due to working overtime. The algorithm underestimates fatigue for participant 6 (KSS = 4.55). This is due to the participant working overtime, with a shift scheduled to finish at 17:00pm, however continues to work until 19:32pm. The alarm sounds 68 minutes after the end of the proposed shift. Participants 3 and 15 were the only participants in this group to score >6 on the KSS. The alarm threshold is triggered in both instances, within $\pm$1 on the KSS.

As previously shown in section 5.2.2, no day shift workers undertaking a normal shift length reported $\geq$6 on the KSS scale. The algorithm considers the average circadian rhythm, time on shift and RMSSD. As such, for day shift workers, two of the three variables increase with time (with respect to breaks taken), whilst the RMSSD can delay the onset of the alarm depending on the individual's activity levels. Generally, towards the end of most participants' shifts, their RMSSD value increases with a reduced HR, inevitably due to relaxing at the end of the shift.

6.4.2 Interpretation of L (Eqn 10)

With Eqn 10, user input is not necessary. It can be shown using Table 6.6, that both Eqn 9 and 10 are equally accurate at notifying the user at an appropriate time, when they are likely to be exhibiting signs of fatigue.

Table 6.6: KSS score when each participant's *L* parameter would trigger the alarm (considering total elapsed time on break) using Eqn 10.

Participant	L	KSS	Time into shift (minutes)	Total elapsed time on break (minutes)	Percentage of shift before alarm
2	N/A	N/A	N/A	N/A	N/A
3	8.17	5.60	504	N/A	41.96
6	8.15	4.55	547	N/A	114.20
12	N/A	N/A	N/A	N/A	N/A
14	N/A	N/A	N/A	N/A	N/A
15	7.58	5.37	721	N/A	85.12

6.4.3 Interpretation of H,L and S for Power Spectral density (PSD) (Eqn 10)

As aforementioned, the H parameter is utilised to outline the dominant sympathetic or parasympathetic nervous system (Figure 4.17), to be computed into the matrix. The L parameter may be able to provide insight into the noise situated within the ECG or concurrent Tachogram, however steps were taken to minimise the noise, by introducing thresholds between R-peaks on the ECG trace and minimal heights of the R-peaks within the signal. The S parameter is able to reflect the LF/HF ratio as shown with scatter graphs (Figures 6.5a and 6.5b). Figure 6.5a and b show very similar outputs, which may be able to predict the LF/HF ratio using chromatics, however with an S parameter normalised between 0 and 1. Whilst the figures look similar, the spread (S) of the data within the PSD graphs does not necessarily affect the rate of fatigue accumulation. Future research should look at the association of the L and S parameters in more detail, as chromatics may be a useful indicator of fatigue identification, with a larger sample size to remove any noise which can be depicted in the figures.

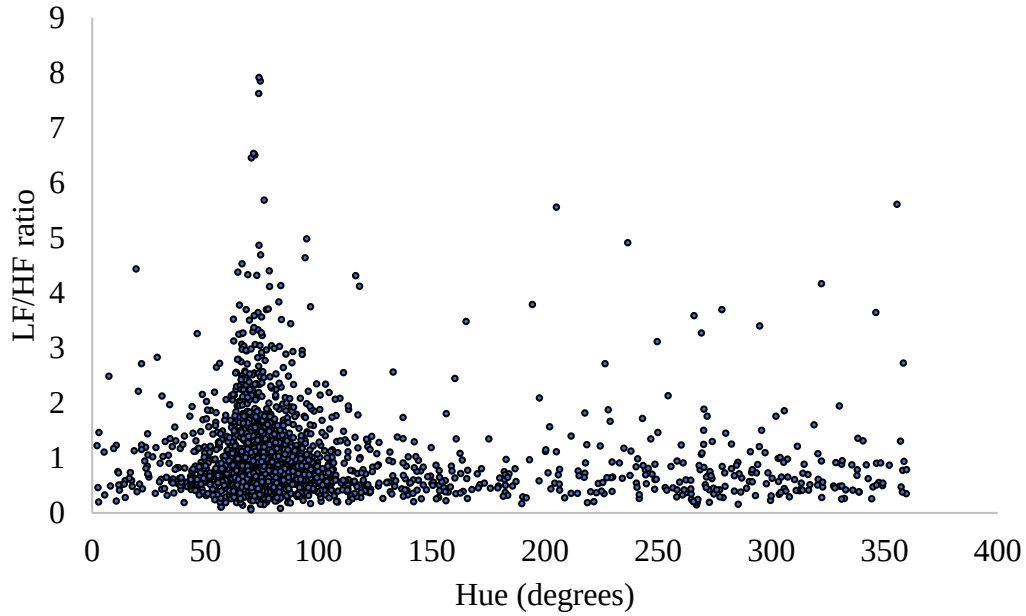


Figure 6.5a: Visual relationship between LF/HF ratio and chromatic parameter Hue.

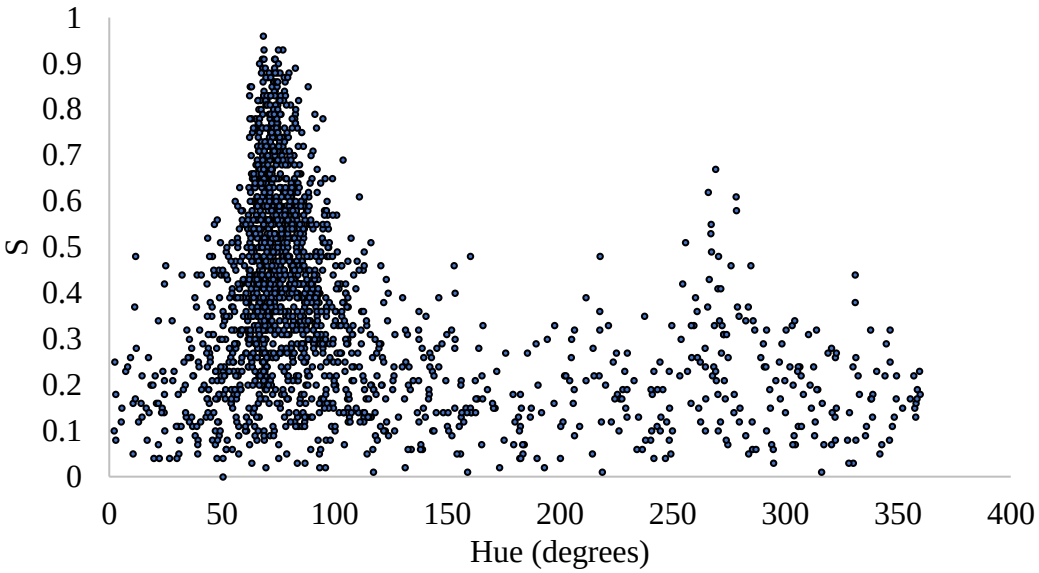


Figure 6.5b: Visual relationship between chromatic parameter Saturation and Hue.

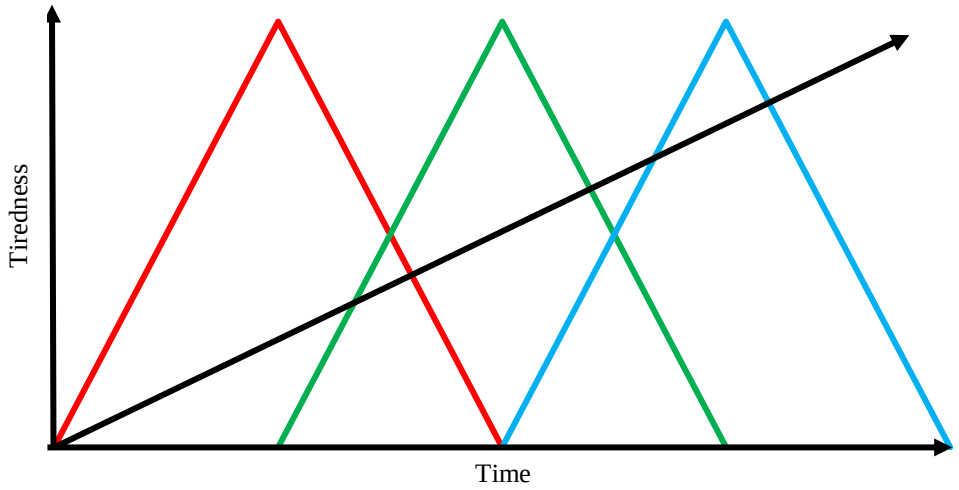
6.5 Chromatic filtering

For the chromatic filtering, time stepping and filter size were optimised for the maximal performance of the algorithm. For all parameters, the filter width spanned five minutes to identify the trend of tiredness and nine minutes for all three filters. Similarly, time stepping occurred every minute for all parameters. This filter size was found to cause the most significant correlation between KSS and *L*. This yields a sampling rate of one data point per minute. An increased filter width increases the computational output of the algorithm and smooths out the *L* value. However, the additional computational cost is not worth the extra time required for the alarm to be sounded. For example, increasing the filters' width to 10 minutes would require a total overlap of 19 minutes. When observing the participants in real-time, 19 minutes is a longer time before sounding an alarm to predict the onset of fatigue. With this technique, a chromatic output can be processed for each time step of the experimental data. Trends are identifiable over time as to the fatigue level of the participant in a visual manner.

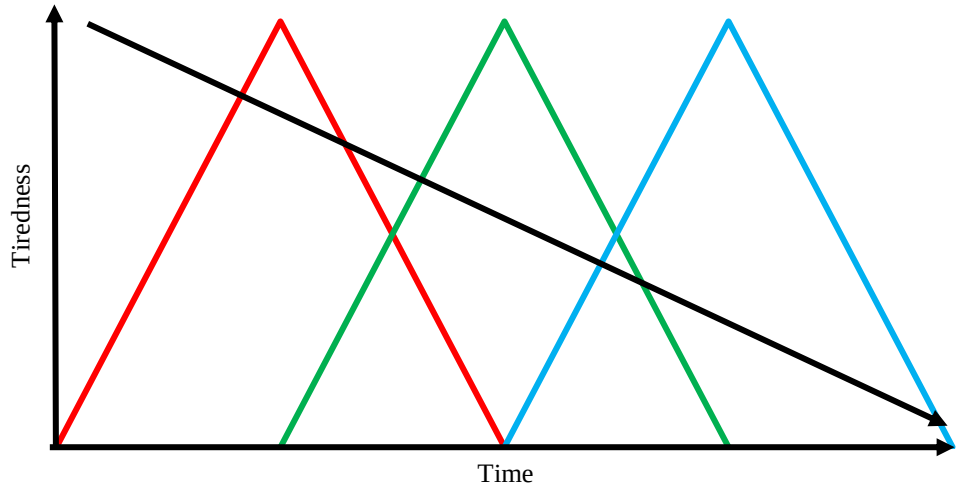
6.5.1 Chromatic visual representation

Figures 6.6 are different signals affecting the proposed advisory system and, subsequently, the position of *H* on the polar plots. The three forms of signals identify the relevant filter strength. Figure 6.6a) shows a positive gradient over time, which results in a dominant blue filter. Figure 6.6b) shows a negative gradient, resulting in a dominant red filter. Figure 6.6c) shows the gradient increases to a peak and then decreases after that, showing that tiredness reaches a peak for the parameter in question.

a)



b)



c)

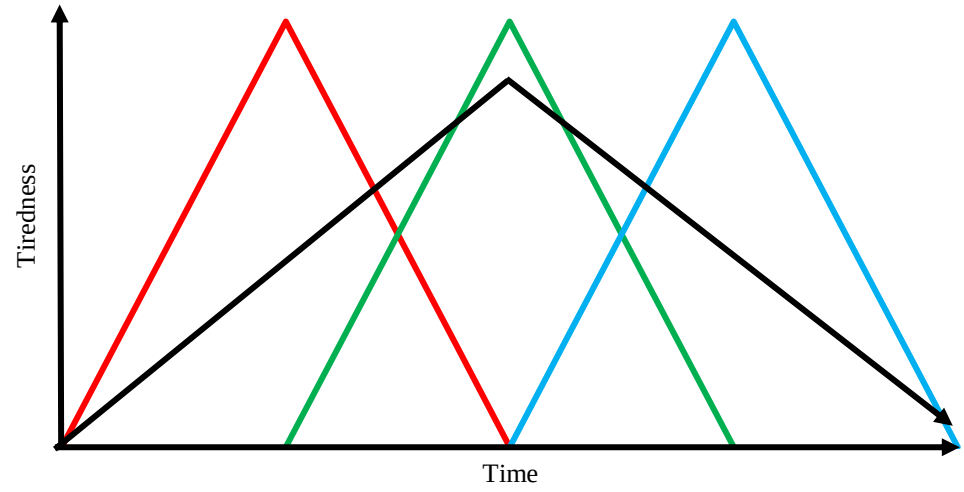


Figure 6.6: Chromatic signal variations (a) Increasing fatigue (b) decreasing fatigue (c) fatigue peak.

6.6 Interpretation of *H*

The polar plot identifies the dominant signal on the respective *R*, *G* and *B* filters. With the filter size selection, the chromatic *H* can distinguish trends of tiredness. Figure 6.7 identifies the correspondence between different sectors of the *H* plot and the subsequent linearity of a signal (Figure 6.6).

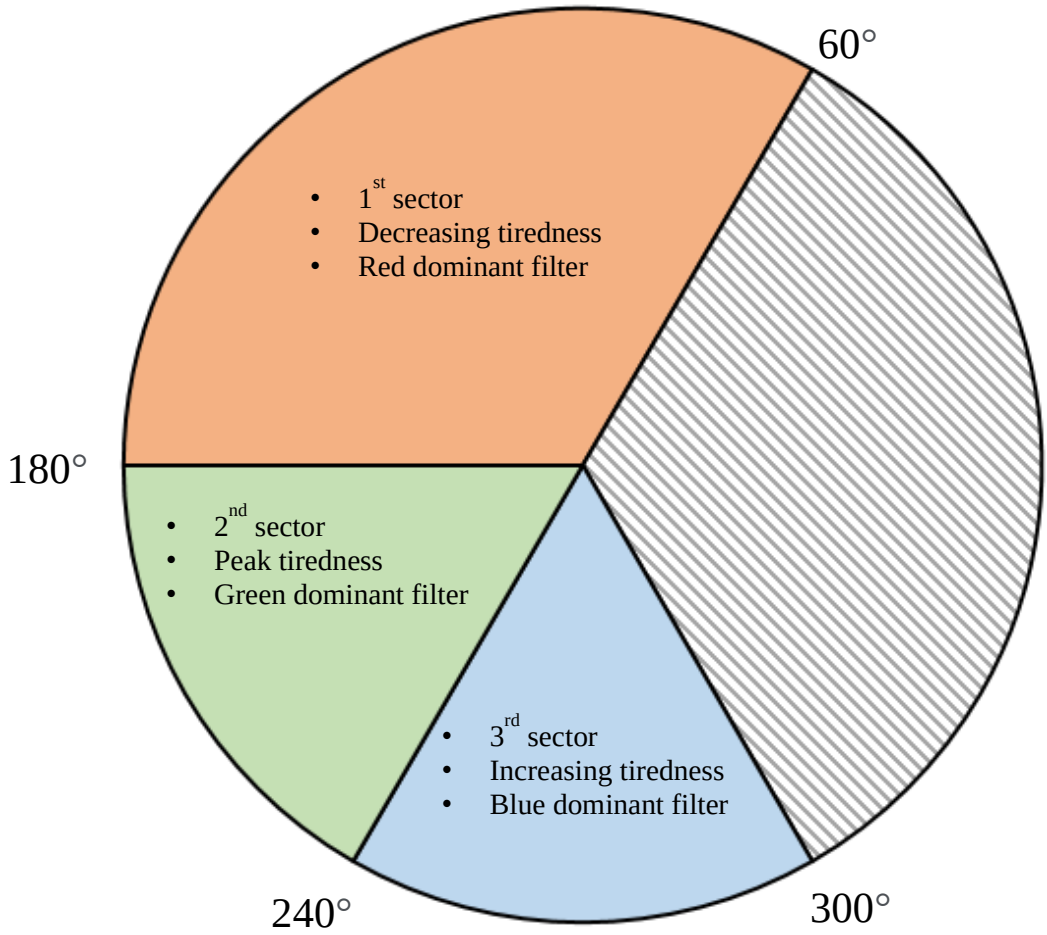


Figure 6.7: Correspondence of sectors of *H* and their respective effects on the signal adopting the signal variations in Figure 6.6.

Applying the chromatic processing to the averages of the parameters RMSSD, Time on Shift and Circadian rhythm outputs the H-L polar plots below. Figure 6.8 represents all data points for all participants when working, whilst Figure 6.9 represents all data points for participants whilst on a break. The H representation is an average of the three parameters.

There is a tendency for the break data to cause an anti-clockwise shift in polar plots due to the reduction in the Time on Shift variable, causing *R* to be the dominant filter. In

Figure 6.8, the participant's *H* data points fluctuate between 120° and 300° degrees whilst working. In Figure 6.9, the *H* data points fluctuate between 60° and 240°. Due to the decreasing nature of the Time on Shift variable whilst on a break, the *R* filter becomes one of the three dominating filters, reducing the angle of *H* closer to 0°.

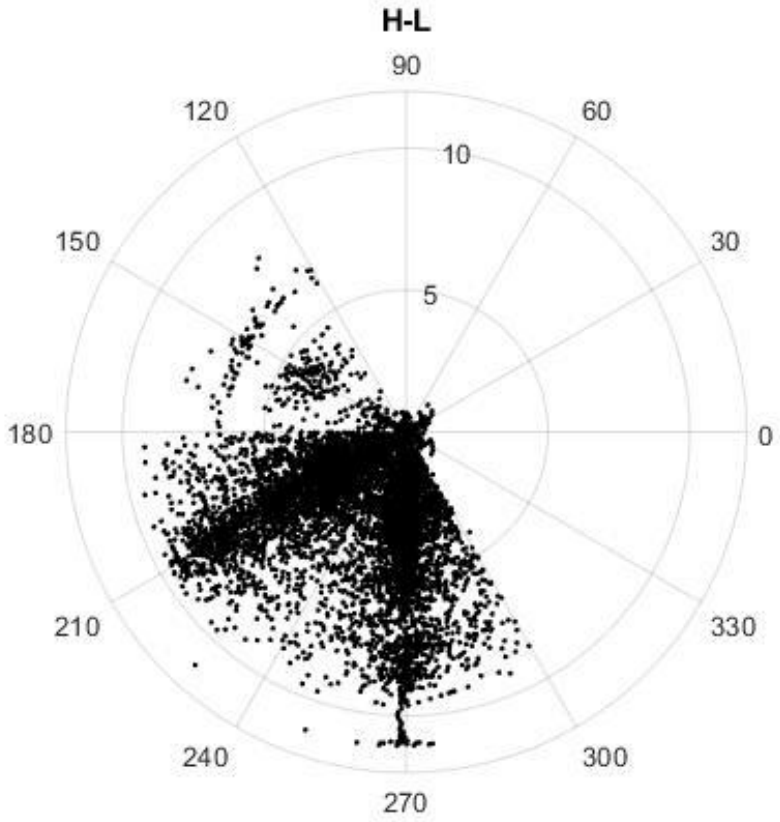


Figure 6.8: All data points for *H-L* plots whilst participants are categorised as working during their shift

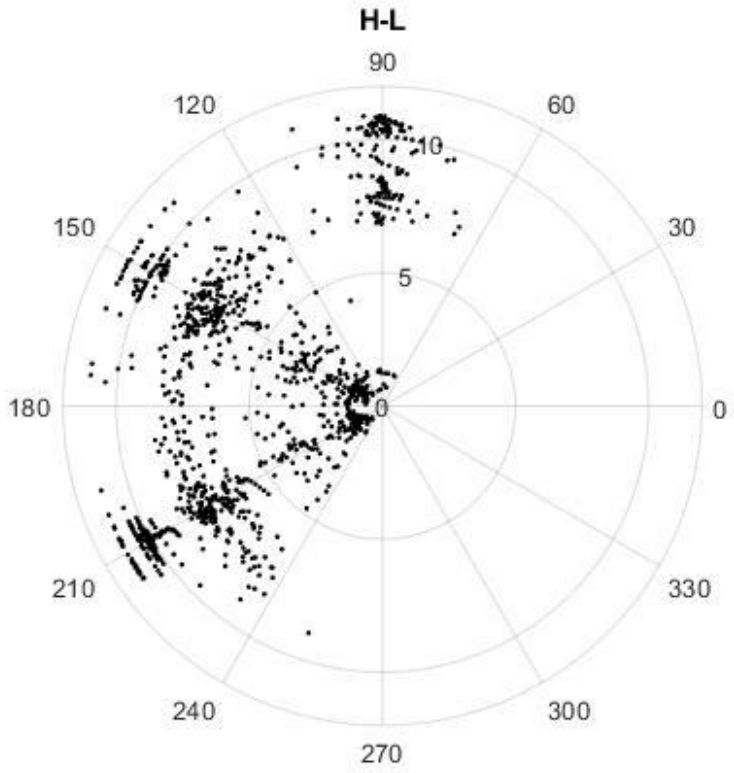


Figure 6.9: All data points for *H-L* plots whilst participants who are categorised as being on a break during their shift.

The relationship between the dominant filters and H plots are identified:

Data points below 240° include at least one *G* or *R* filter, indicating a decline in circadian rhythm, Time on Shift or RMSSD. Figure 6.9 has one outlier for the break, with a *H* value of 251°. This is due to a sharp increase in RMSSD, circadian rhythm and a peak for Time on Shift, which resulted in an outlier. Recording a *H* value less than 240° is generally good and results in a more steady increase in *L*.

A *H* lower than 180° includes at least one *R* and one *G* dominant filter, increasing the delay before the alarm is sounded due to a shallower increase in *L*.

Data points above 240° indicate a maximum of one *G* filter but no *R*, meaning an incline in the variables overall, including a minimum of two *B* filters.

Data points over 270° (Figure 6.8) include three *B* dominant filters. Therefore, the participants in this category are at risk of becoming fatigued at a greater rate, resulting in a steeper increase in *L* and a reduced time before the alarm is triggered.

The Kruskal-Wallis test showed that there was a statistically significant difference in the change in *L* between consecutive values for the three different *H* zones (Figure 6.7) with the exclusion of breaks, $\chi^2(2) = 2994.63$, $p < 0.001$, with a mean rank *L* score of 1928.83 for zone 60-180°, 2602.68 for zone 180-240° and 5292.56 for zone 240-300°. The mean results are displayed in Table 6.7 (below). The trend line shows a gradual increase in *L* steepness with an increase in *H*, displayed in Figure 6.10.

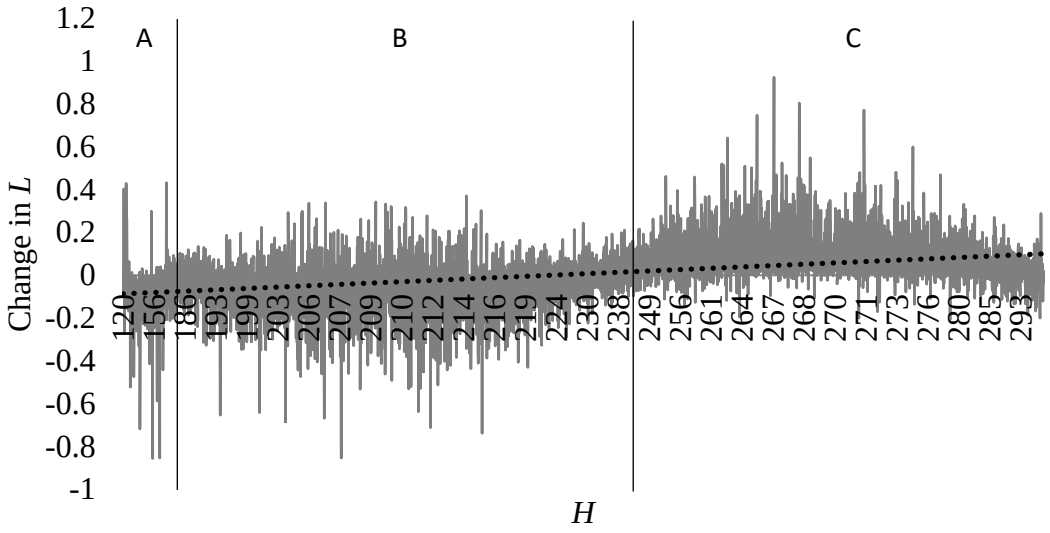


Figure 6.10: Difference in consecutive *L* values whilst working, by *H* zone (A = 1st sector, B = 2nd sector, C = 3rd sector) (Figure 6.7).

Table 6.7: Comparison of mean change in *L* split by sector (Figure 6.7).

Sector	Mean (Standard Deviation)	Sample (n)
1st	-0.08(±0.12)	363
2nd	-0.05(±0.11)	3772
3rd	0.08(±0.1)	3453

Whilst there is an upward trend of increased tiredness in the 3rd sector of Figure 6.7, no useful additional, operational information is produced for diagnosing the onset of workplace fatigue and therefore the *H* parameter's use in the algorithm is not necessary in this application, due to the large variation in the dominant filter as a result of the RMSSD variable.

6.6.1 Application of H

In this application, the Hue (*H*) of a signal does not provide much meaningful data to predict the onset of fatigue. Figure 6.11d contains the average data from Figures 6.11a-c (with respect to the day participant in Figure 6.11a). As can be shown the data from Figure 6.11c causes drastic shifts in the dominant filter, causing large fluctuations.

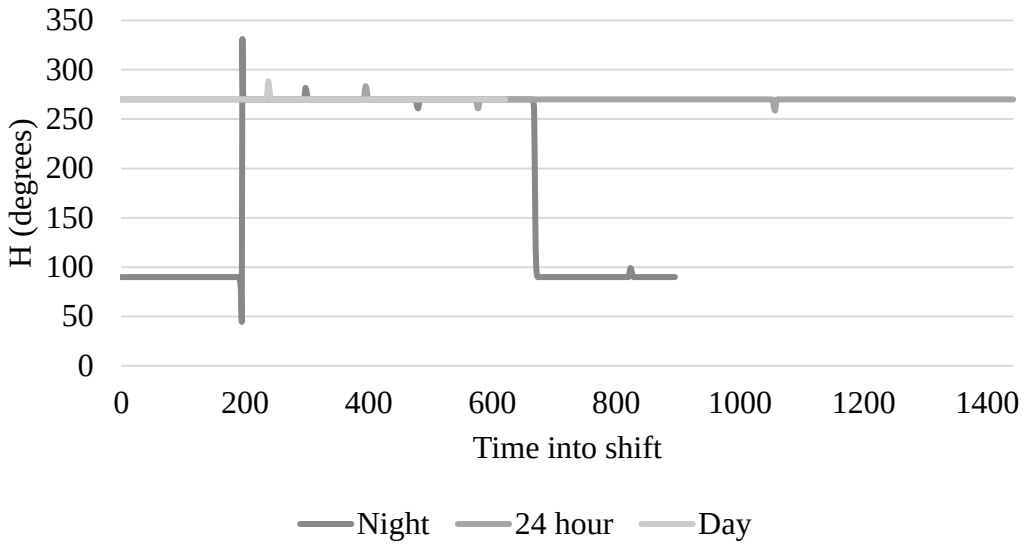


Figure 6.11a: Hue in degrees of predicted KSS with respect to shift type, versus time into shift for a day-shift participant.

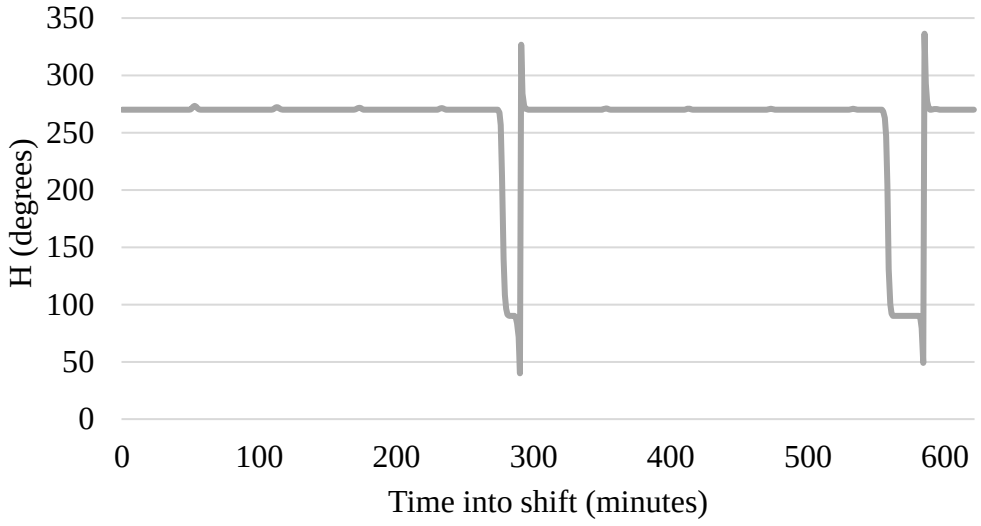


Figure 6.11b: Hue in degrees of Time into shift, versus time into shift for a day-shift participant, showing examples of two breaks being taken that effect the dominant filter.

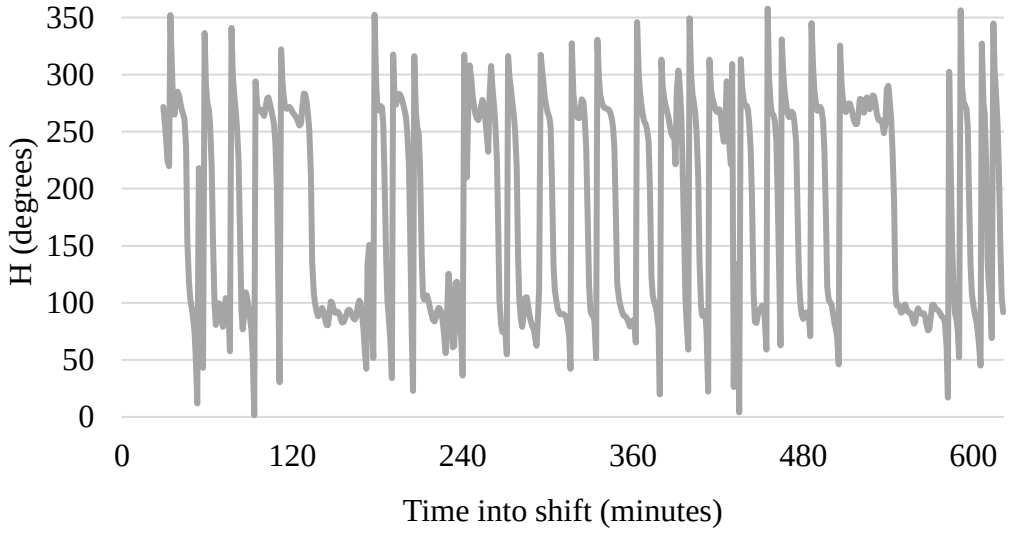


Figure 6.11c: Hue in degrees of RMSSD versus time into shift for a day-shift participant.

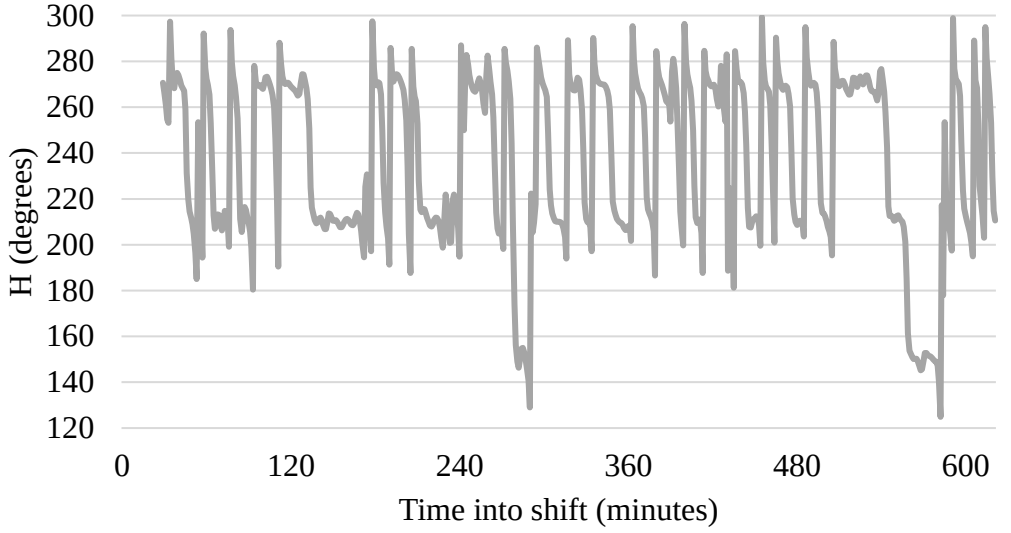


Figure 6.11d: Hue in degrees of combined measures from Figures 5.16a-c versus time into shift for a day-shift participant.

Due to the large variation in the RMSSD measure, the H of the signal in this application does not provide meaningful data for the prediction of workplace fatigue. As shown in Figures 6.8 and 6.9, there is no linearity with the data and as such data presentation in the form of polar plots cannot be executed to provide meaningful information for the prediction of workplace fatigue.

6.6.2 Example of participants' *H-L* polar plot data split by work and break

The following section includes examples of participants' *H-L* plots. The left polar plot depicts the time spent working, and the right polar plot shows the time spent on breaks. The sections of the shift are colour coded to represent the different stages of work:

- Red - first period of work and first break, respectively
- Green – second period of work and second break, respectively
- Blue – third period of work and third break, respectively
- Yellow – fourth period of work and fourth break, respectively

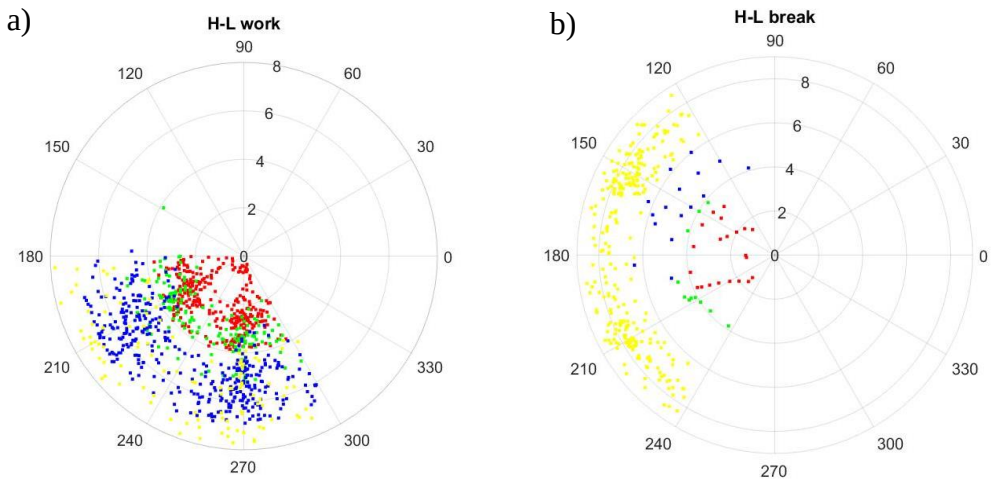


Figure 6.12: Example of *H-L* plots for one participant, split by (a) work and (b) break periods, where the colour depicts each iteration of work and break during a shift.

Figure 6.12 show an example of a 24-hour shift in *H-L* polar plots, where each data point represents one minute. It is shown that a clockwise shift of 60° occurs whilst on break due to the *R* dominant filter. This pattern is the same for all participants. The radius of the data points increases over time due to the Time on Shift and Circadian rhythm variables. The radius of *L* is slightly decreased on break, compared to working, due to the decrease in the Time on Shift variable.

A crossover is observed between the blue and yellow periods whilst working, where the break reduced the *L* value back below the advisory alarm threshold for some time. In this example, the participant began their break with an *L* value = 6.54 and, after a

15-minute break, was able to reduce their *L* value = 4.18 due to a reduction in RMSSD and Time on Shift variables. However, as aforementioned, the *H* parameter does not provide meaningful data in order to predict the onset of workplace fatigue.

6.7 Interpretation of *S*

The chromatic *S* measures the spread of the signal. Signals with little spread result in a more considerable *S* value, and vice versa. The RMSSD parameter fluctuates every minute, resulting in significant changes in the spread of the signal that cannot be used to predict tiredness. Due to the high frequency of time stepping, the *S* parameter also does not yield any valuable information for the proposed advisory system in this application.

Figure 6.13 include all *H-S* plots from all participants whilst working and on break, respectively. The *S* parameter contains negative numbers due to the RMSSD minimum and maximum parameters, which affect the weighting of the *L* parameter. For example, an RMSSD value of 24, when the minimum parameter for RMSSD is 25, results in a normalised value of -0.2. As such, it is difficult to read the results of *S* using polar plots because the negative values mirror the reading of *H*.

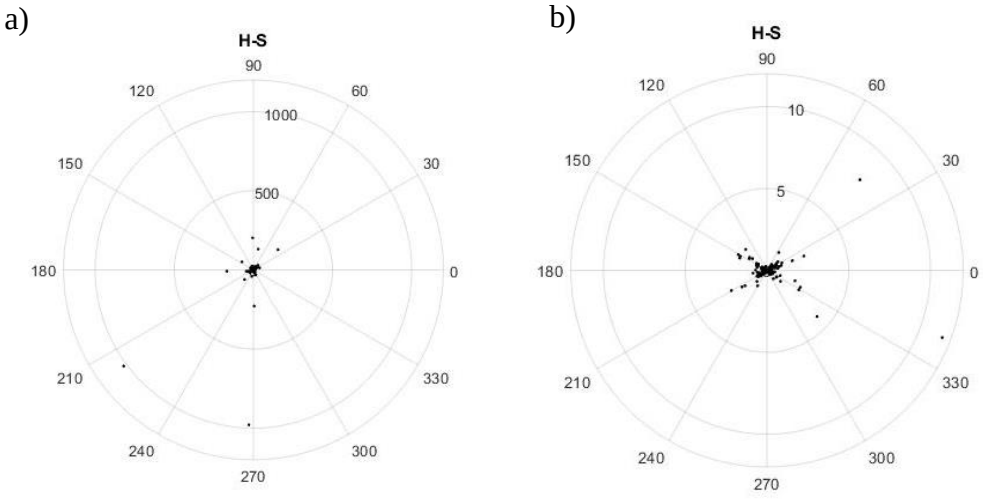


Figure 6.13: All data points for *H-S* plots (a) whilst working (b) whilst on break across all participants.

No relationship was observed between the *S* value and Heart Rate (HR), Heart rate variability (RMSSD) or acceleration (VM) for increasing the KSS score, whereby a lower *S* value did not increase the KSS scores. This may have been useful in instances where participants were stationary for long periods of time and therefore had relatively stable measures of HR and VM.

6.7.1 Application of S

Figures 6.14a-e below show the *S* components of the variables utilised in the algorithm and their respective raw values. Predicted KSS for each shift type is presented in Figures 6.14a-c, the Time into shift is presented in Figure 6.14d and the RMSSD is presented in Figure 6.14e. The *S* component displays spikes in the spread of a signal when there is deviation in the raw values. The data in Figures 6.14a-e does not report meaningful data for the identification of workplace fatigue in this application as the data is held within pre-determined, normalised parameters. This causes large spikes to be present in instances where the raw data equals the minimum or maximum parameter, such as the spikes identified in Figures 6.14a-b.

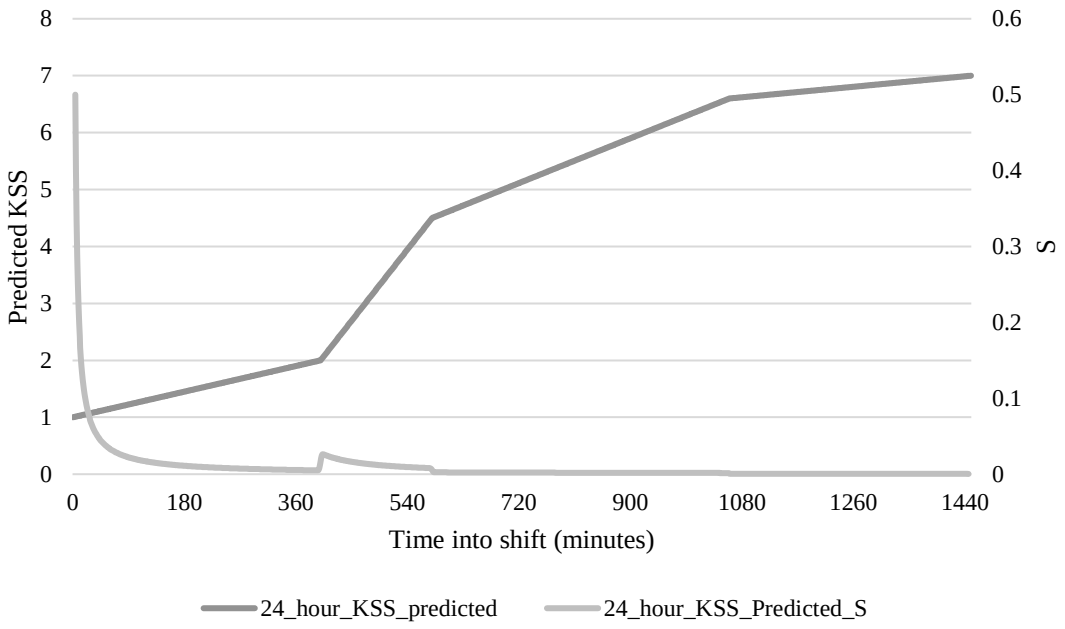


Figure 6.14a: Comparison of the spread of the predicted KSS scores for 24 hour shift workers against the raw data for predicted KSS.

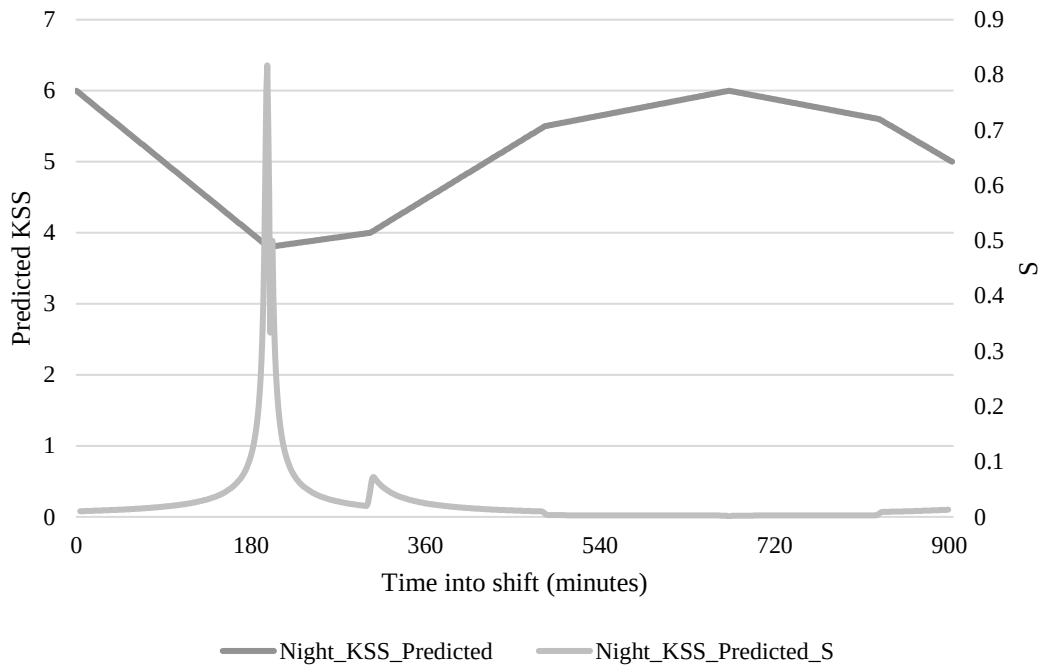


Figure 6.14b: Comparison of the spread of the predicted KSS scores for night shift workers against the raw data for predicted KSS.

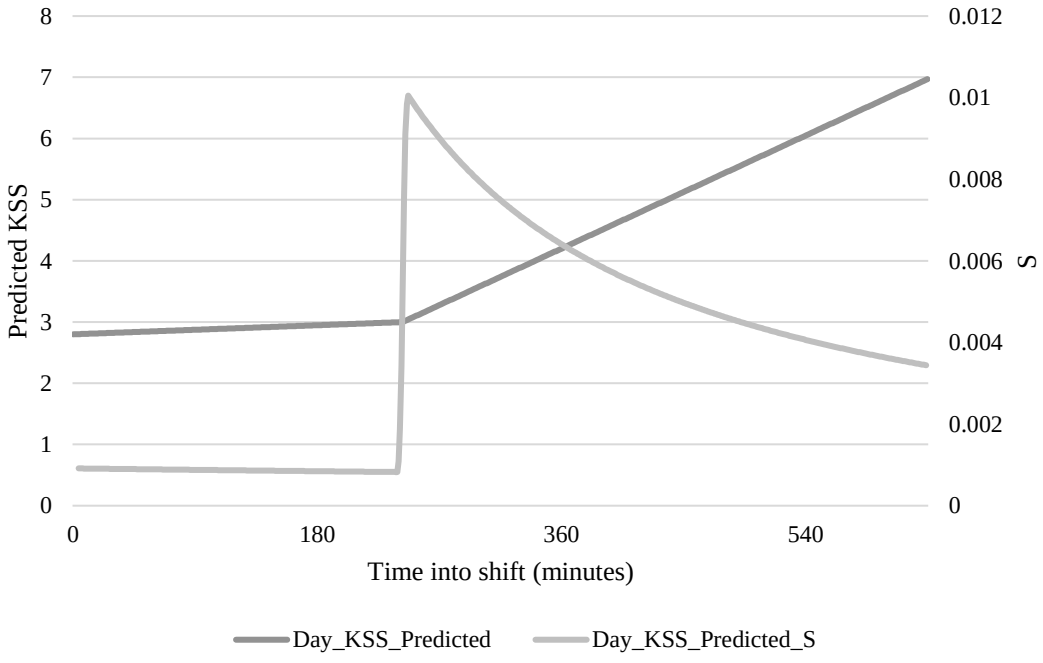


Figure 6.14c: Comparison of the spread of the predicted KSS scores for day shift workers against the raw data for predicted KSS.

The *S* data in Figure 6.14d offers more variation depending on the amount and length of breaks taken during the shift. In this example, a large spike is shown towards the end of the shift, as the length of the break returns the Time on shift score back towards the minimum parameter set for normalisation of the variable.

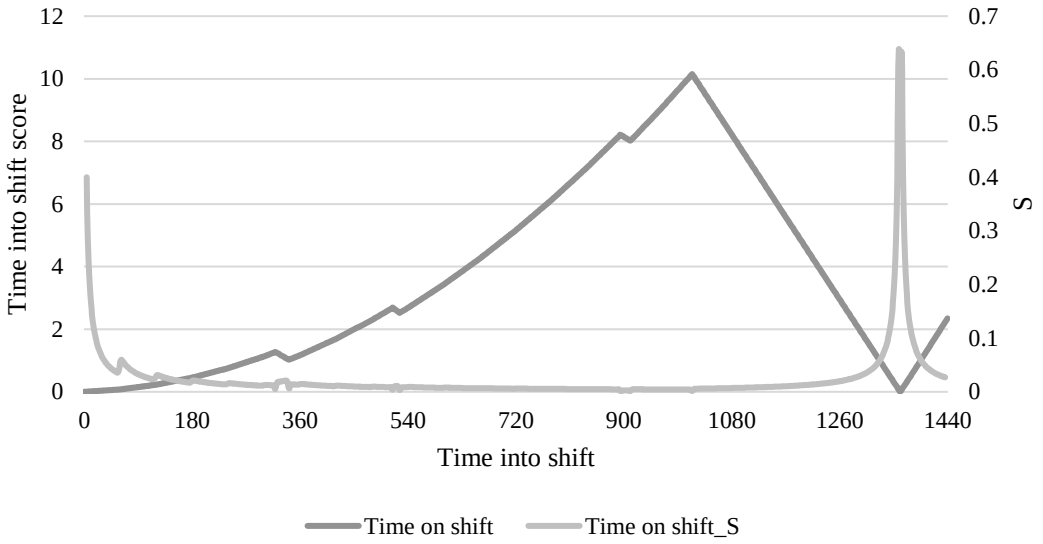


Figure 6.14d: Comparison of the spread of the Time on shift scores for one participant working a 24 hour shift participant against the raw data for Time on shift score using Eqn 9 for the deduction depending on time on break.

Finally, Figure 6.14e offers the most variance, however *S* in this case cannot determine tiredness. A low *S* value is a product of a relatively stable Heart Rate (HR) over time, however the HR may be stable at high or low beats per minute. The analysis in this study identifies high RMSSD measures to be a predictor of KSS scores, therefore the spread of the signal cannot determine the level of tiredness as a product of time on shift or demanding work.

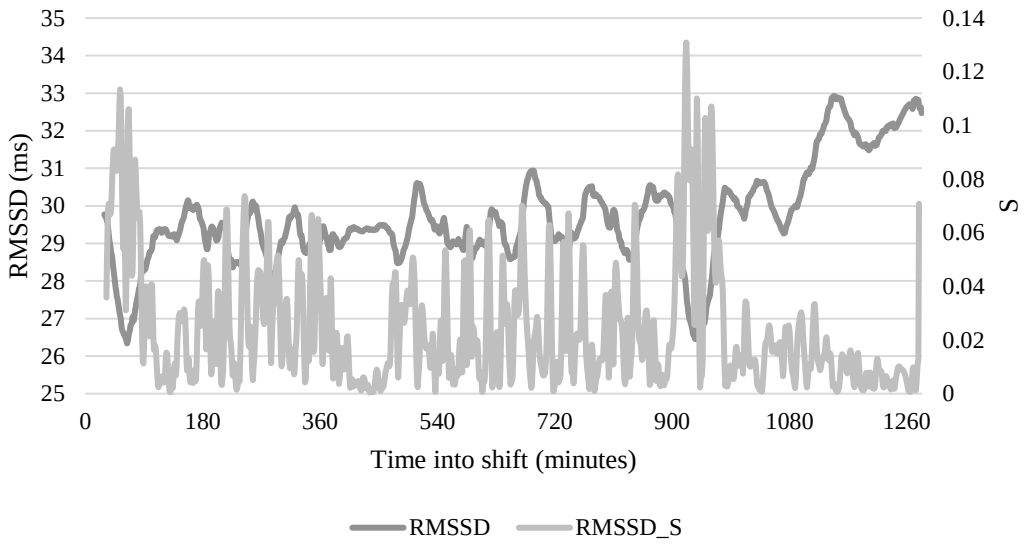


Figure 6.14e: Comparison of the spread of the RMSSD for one participant working a 24 hour shift against the raw data for RMSSD.

6.7.2 Example of participants' *H-S* data split by work and break

Figures 6.15 show the *H-S* polar plots for the same participant depicted in Figure 6.12. Across all participants, the *S* is lower on average whilst on a break when compared to working. The yellow markers for the break polar plot show very little spread over time due to the participant taking the opportunity to sleep whilst on shift. The RMSSD output is relatively stable over time due to the participant's lack of physical or mental stressors observed whilst awake. As such, the *H-S* polar plot may be used to observe the fluctuation in RMSSD, as circadian rhythm and Time on Shift follow linear increases or decrease over time. However, this does not necessarily predict the onset of fatigue in the workplace but the stability of one's sympathetic nervous system.

The following section includes examples of participants' *H-S* plots. The left polar plot depicts the time spent working, and the right polar plot shows the time spent on breaks. The sections of the shift are colour coded to represent the different stages of work:

- Red - first period of work and first break, respectively
- Green – second period of work and second break, respectively
- Blue – third period of work and third break, respectively
- Yellow – fourth period of work and fourth break, respectively

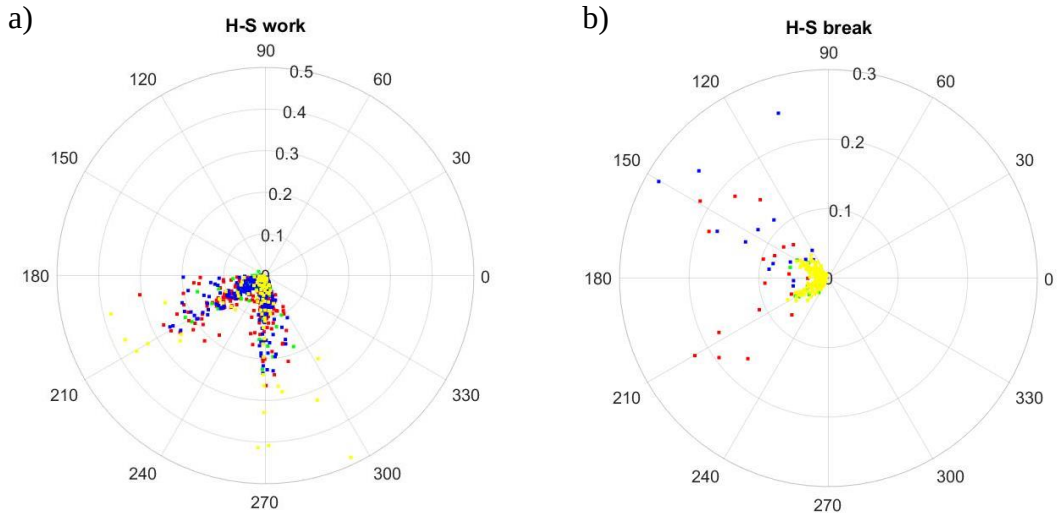


Figure 6.15: Example of one participants *H-S* plots, split by (a) break and (b) work period, where the colours denote each iteration of break and work.

6.8 Conclusion

This chapter described the application of chromatics from the experimental results derived in chapter 4 to create the building blocks of a novel algorithm capable of identifying the onset of fatigue in surgeons. The algorithm is simple, consisting of one physiological parameter and two physical systems to produce >94% accuracy within $\pm$1.96 standard deviations between the KSS and the *L* parameter. This allows for easy administration of such a product in the workplace, where it is unlikely to cause discomfort or affect the work of surgeons. The data for RMSSD is smoothed with a moving average so that there are less sharp spikes in the average *L* parameter, leading to false or frequent alarms.

Breaks are shown to reduce the KSS of an individual; they are accounted for in the algorithm. The breaks reduce the Time on Shift variable, which rises with increased time on shift (Koh et al., 2007). Therefore, taking breaks delays the onset of the alarm threshold being triggered. Shorter and more frequent breaks reduce the Time on Shift variable more effectively due to the rise of Time on Shift, which the administration of a break linearly reduces.

Chromatics can be used to visualise the data using the *H* parameter, where an angle closer to 0° indicates less fatigue. The *L* parameter can define the magnitude of fatigue. As the shift progresses, the magnitude of *L* increases. Once the *L* parameter reaches a defined magnitude, an alarm may be triggered to advise the user to take a strategic break before the participant is likely to become excessively tired. The visualisation of the data can show, in real-time, how fatigued a participant is by observing the trend of their data. Data points over 270° include three *B* dominant filters. Therefore, participants in this category risk becoming fatigued at a greater rate, resulting in a steeper increase in *L* and a reduction in time before the alarm is triggered.

The algorithm provides a holistic means for identifying fatigue in the workplace for surgeons (shift workers) and, as such, acts as an advisory system that can be ignored upon the recipient's request. If the alarm is silenced, the following alarm may sound when the participants predicted tiredness increases by one on the KSS (Table 6.2) and a break had not been taken.

Chapter 7

Application of break prompts using advisory system

7.1 Introduction

The following chapter show group B results for both Eqn 9 and Eqn 10, as to how well they predict the onset of fatigue in the workplace. The overall correlation, including Bland-Altman plots, will be displayed for the reproducibility of the algorithm and it's effects at predicting KSS. This chapter will apply the concepts of the algorithm into practice, as well as demonstrating how multiple alarms may be triggered to suggest break prompts to surgeons on shift depending on the algorithms described in Chapter 5 (Eqn 9 + 10), using the following rules for Eqn 9:

1) Alarm does not sound within 60 minutes of a break

In order to prevent alarms occurring after participants have scheduled a break, the alarm will not sound shortly after a break. The break may have been used as an opportunity to consume caffeine, which takes time to enter the blood stream. Simultaneously, the user may be continuing on with the same task as before the break, however the break gave the user opportunity to disconnect from the task for a short-time so that they are more alert on the task afterwards (Nejati et al., 2016).

2) Threshold for alarm is increased relative to the time on break

The time on break is considered, as breaks are administered as opportunity to distract from the task at hand, if used appropriately. As such, the threshold is extended so that those who take breaks do not reach the threshold at a similar time to those who do not take a break, dependant on heart rate variability (HRV) scores. The time on shift variable is also reduced, further delaying the onset of the alarm.

7.2 Analysis of Eqn 9

Table 7.1 shows the instances where the alarm sounds for the second group of participants, with reference to Table 6.2 and the aforementioned Eqn 9 criteria.

Table 7.1: The analysis of group B for Eqn 9, showing instances where the alarm sounded.

Participant	L	KSS	Time into shift (minutes)	Total elapsed time on break (minutes)	Percentage of shift before alarm
1	N/A	N/A	N/A	310	N/A
8	N/A	N/A	N/A	105	N/A
9	N/A	N/A	N/A	30	N/A
10	N/A	N/A	N/A	0	N/A
11	N/A	N/A	N/A	25	N/A
13	8.49	3.33	622	30	75.95

In this set of data, Participants 1 and 8 exceed the KSS >6 threshold, but the alarm does not sound. For participant 1 (Figure 7.1), they log a break from 00:45am until the end of the shift (post 08:00am). Therefore, as they are administering ‘sleep opportunity’, the alarm does not sound. As for participant 8 (Figure 7.2), the alarm does not sound also due to the time spent on a break. However, the time on the break was fairly early on into the shift, when it would have been more appropriate to administer break opportunity later on into the shift, where more benefit would have been sought. The alarm triggers for participant 13, however the KSS is over-estimated in this example (Figure 7.3).

The line graphs in this chapter (e.g. Figure 7.1, below) show the signal magnitude (L) of the variables used in the algorithm over time, against the actual KSS scores. The alarm threshold is depicted using the red line and average is the average L of all the signal magnitudes. This can be used to track workplace fatigue in real time.

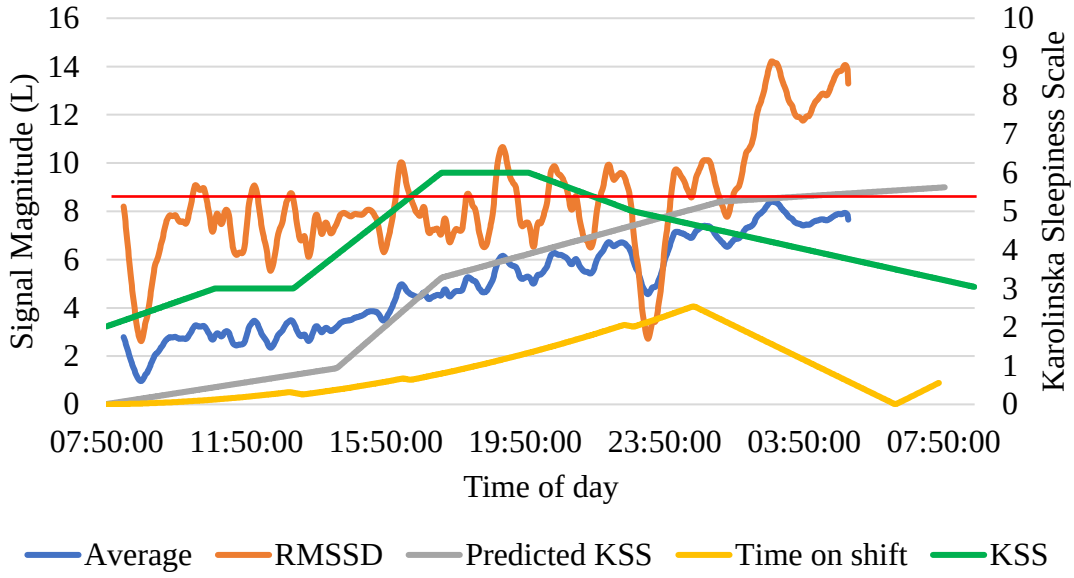


Figure 7.1: The alarm not triggering for participant 1, due to the sleep opportunity taken at the end of the break.

Participant 1 (Figure 7.1) logs a 'good' sleep quality prior to the shift, which is why the average signal magnitude pictures is relatively low. The alarm would have triggered in this example, had the participant not logged sleep opportunity at 00:45am.

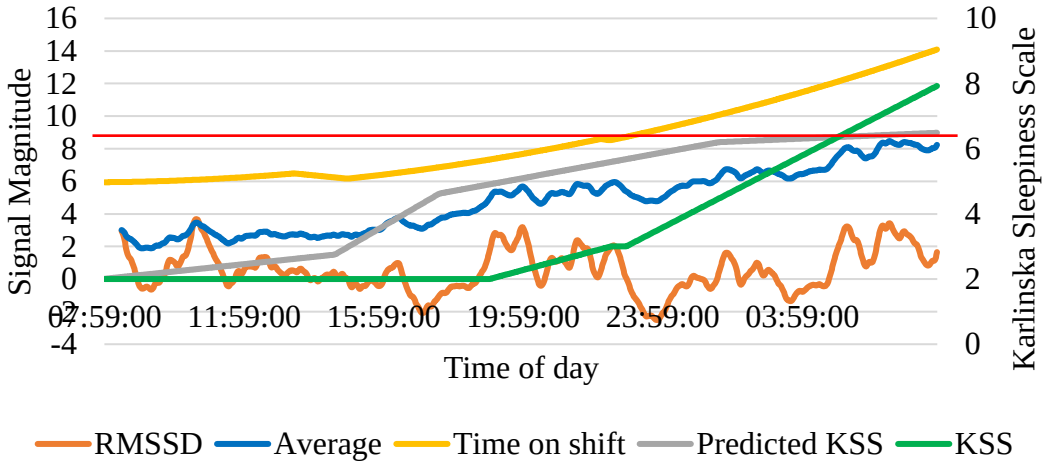


Figure 7.2: The alarm not triggering for participant 8, due to the increase in the alarm threshold, as breaks have been taken during shift, as well as low RMSSD measures.

Participant 8 (Figure 7.2) logs a poor sleep at the beginning of the shift, which explains the large increase on the y-axis for the time on shift parameter. Despite this, the alarm does not flag as the participants RMSSD remains low throughout the shift. A consideration with this algorithm, is that it assumes KSS increases and decreases linearly, which may not be accurate. Future research needs to look at the effects of KSS with shorter time-points between questionnaires, as this participant may have been feeling alert for the majority of the shift, but when finally given time to rest at the end of the shift (filling in another questionnaire), they realise how tired they are. The participant did not log a KSS score for 8 hours prior to the final one, therefore it is uncertain how they were feeling at around 4am, when a KSS should have been logged. Nonetheless, without this information, it is assumed that the KSS increased gradually over time.

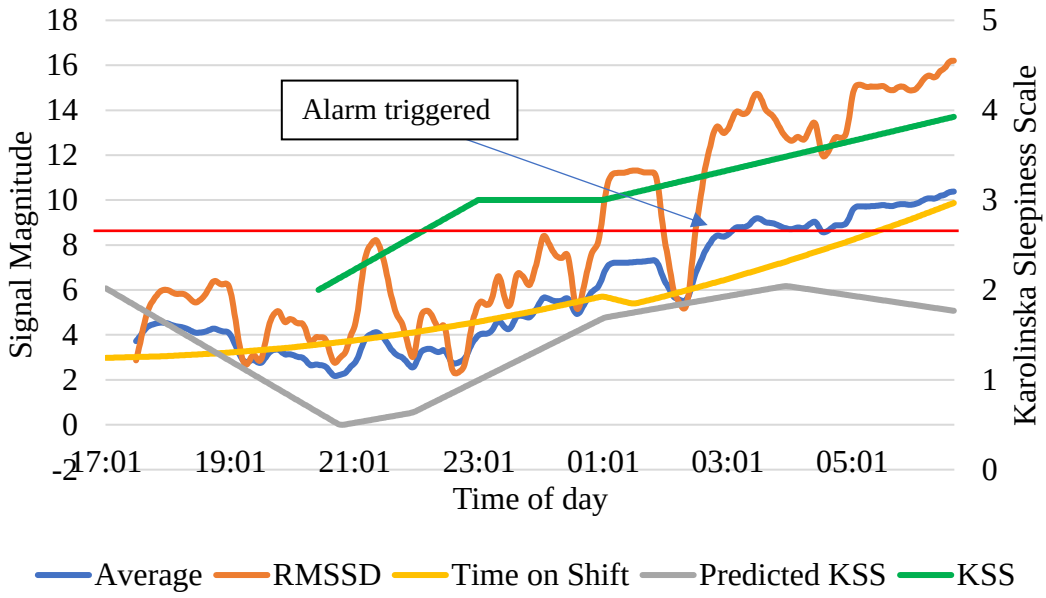


Figure 7.3: The alarm triggering for participant 13, due to very high RMSSD measures during the latter half of the shift, as well as it being a night shift with only one 30 minute logged break.

Participant 13 (Figure 7.3) demonstrates an example, where with an 'average' logged sleep, high RMSSD measures and it being a night shift, where the predicted KSS is expected to be around 6, that the alarm sounds. In this instance, the alarm over-estimates the actual KSS. It is expected, that for the participants RMSSD to be so high, that they must be engaging in very little activity, presumably monotonous desk work. Whilst the alarm over-estimates the actual KSS, the participant is made aware that this time of day

with a low HR are warning signs that errors may be made due to a high-fatigued state. In the example, the alarm threshold is: 8.14 + 0.3394 = 8.48. The alarm is triggered at 03:06am (see blue circle on Figure 7.3).

7.3 Analysis of Eqn 10

Table 7.2 shows when the alarms are triggered using the Matrix, which includes analysis of rest opportunities by the participant, with reference to vector magnitude (VM) and power spectral density (PSD), using chromatic H to determine the state of the autonomous nervous system (ANS). The Matrix is able to categorise medium or high fatigue events, as well as restorative events (of at least 5-minutes).

Table 7.2: The analysis of group B for Eqn 10, showing instances where the alarm sounded.

Participant	L	KSS	Time into shift (minutes)	Total elapsed time on break (minutes)	Percentage of shift before alarm
1	8.15	4.32	1116	N/A	86.92
8	8.16	6.41	1268	N/A	88.00
9	N/A	N/A	N/A	N/A	N/A
10	N/A	N/A	N/A	N/A	N/A
11	N/A	N/A	N/A	N/A	N/A
13	8.17	3.65	721	N/A	86.94

In comparison to Eqn 9, the alarm threshold is triggered for participants 1 and 8. In this equation, the time on logged breaks is not considered. Figures 7.4, 7.5 and 7.6 show the comparison between participants 1,8 and 13, as to when the alarm is triggered. For participants 9, 10 and 11, the KSS does not $\geq$6, therefore the alarm is not necessary and works appropriately.

Figure 7.4 demonstrates the alarm triggering when logged breaks are not considered, rather ‘true’ restorative periods at work being monitored. The Time on shift component has a different shape compared with Figure 7.1. In this example, whilst the participant had administered a break for ‘sleep opportunity’ it is more apparent, that the participant

does not meet the criteria which is representative of restoration, either because the participant is moving or the participants PSD is not reflective of parasympathetic dominance (G or B filters). Therefore, the participant is still in a state where stressors are influencing the individual. The alarm is indicated by the blue circle on Figure 7.4, it triggers at 00:54am.

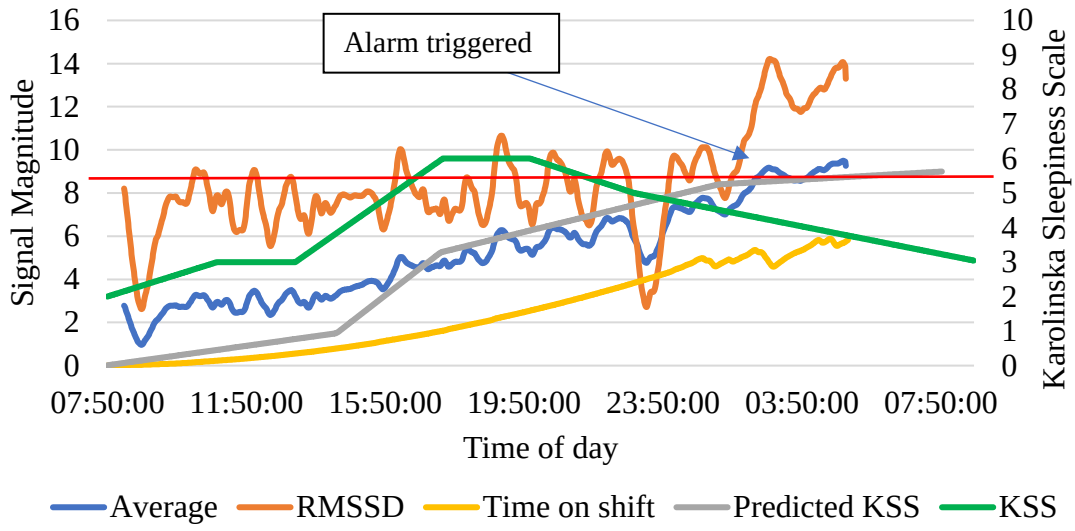


Figure 7.4: The alarm triggering for participant 1, due to high RMSSD and high predicted KSS, with little restorative break being taken.

For participant 8, the alarm triggers at 05:14am. The participant has no restorative breaks during the shift, therefore the time on shift parameter rises. Due to the high offset as a result of logging poor sleep quality, the participant is correctly identified as being fatigued in this scenario, despite the low RMSSD measures, proving that the algorithm can account for a variety of factors which may otherwise seem counterintuitive.

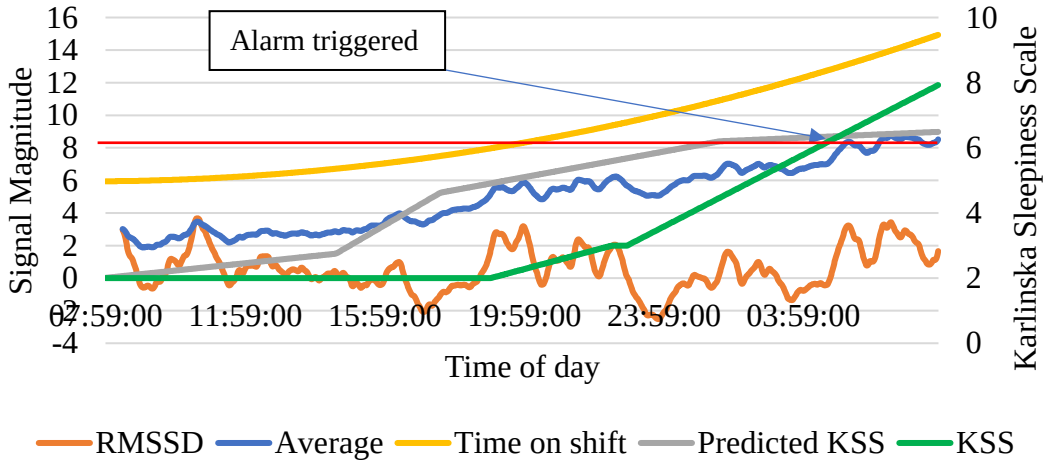


Figure 7.5: The alarm triggering for participant 8, due to poor sleep score and high predicted KSS, with no restorative break being taken.

For participant 13 with Eqn 10 (Figure 7.6), the alarm also triggers, due to no restorative breaks being taken and a large RMSSD value. The alarm is triggered at 02:49am. Had the participant recorded good sleep, the alarm would have sounded at 03:22am. In this example, the frequency domain (PSD) and time domain (RMSSD) HRV measures appear to be contradicting each other. The time on shift decline is indicative of both PSD and vector magnitude (VM). It may be the participant is seldom stationary and therefore a 'true' restorative period is not marked. This means that the participant is engaging in monotonous tasks, where activity is being reported, however in low quantities.

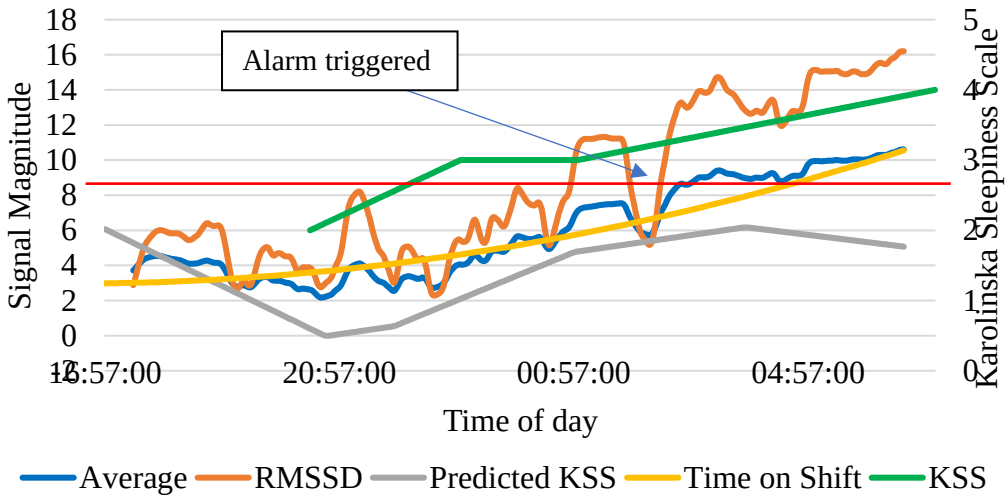


Figure 7.6: The alarm triggering for participant 13, due to average sleep score and high predicted KSS, with no restorative break being taken.

On the other hand, it is possible to have a high RMSSD and a R dominant filters at the same time, as these two measures capture different aspects of heart rate variability (HRV). RMSSD is a time-domain measure of HRV that primarily reflects parasympathetic activity, while the R dominant filter (indicative of a high LF/HF ratio) is a spectral analysis technique that assesses the relative balance between sympathetic and parasympathetic activity. While parasympathetic activity is typically associated with higher RMSSD values, sympathetic activity can also contribute to LF/HF ratio.

For example, during exercise or stress, sympathetic nervous system activity increases, leading to an increase in LF/HF ratio. At the same time, there may be a short-term increase in heart rate variability due to parasympathetic withdrawal and subsequent

reactivation, resulting in a temporary increase in RMSSD. Therefore, it is possible to have a situation where the R filter indicates sympathetic dominance, while RMSSD remains relatively high due to a short-term increase in parasympathetic activity.

7.4 Comparison of Eqn 9 and Eqn 10

The key difference between these equations, is that Eqn 9 requires user input in order to alert the algorithm of when a break is being taken. This can be a benefit as to one's interpretation of the restorative nature of breaks. As described by Horne and Reyner (1996), caffeine and naps which are less than 15 minutes in duration can have excellent restorative effects to prologue safe driving. However, strategic napping is controversial in health care workplaces (Nejati et al., 2016) as health care leaders remain sceptical about the effectiveness of these programmes, despite research indicating otherwise (Arora et al., 2006).

Breaks are considered an opportunity to take on a different mind-set. Meditation or napping alone is not the only scenario in which the restorative effects of a break can be known, however the literature into workplace fatigue and the restorative effects on reducing errors or improving quality of patient care in surgeons is seldom. Using Eqn 9 allows the algorithm to demonstrate that an opportunity to rest is being taken, however the user may over/under estimate the time on and off their break, or even forget entirely, and therefore is subject to user error. It also requires the user to log this information when they may be busy, or called off their break abruptly if there is an emergency in the workplace. As demonstrated in Figure 5.3, there are plenty of instances where participants are taking breaks but they do not feel any more alert afterwards. This may be because they are already alert at the start of their breaks and therefore taking breaks at the wrong time, or perhaps they are not taking an efficient break period.

Eqn 10 does not require user input. It mitigates the possibility of user error and focuses on objective data sets as to the quality of the restorative periods being taken. However, there is a drawback to this equation, if there is not an appropriate place for surgeons to benefit from the restorative effects of sleep or to be in an environment with no stressors.

A break room is the obvious place to go for a break, however it may be full of co-workers, all of who have had equally stressful days. Nejati et al. (2016) describe in detail how nature, the outdoors or windowed offices can have significant impact into improving restorative periods in the workplace.

Table 7.3 is a comparison of the accuracy between Eqn 9 and Eqn 10, and how well they predict fatigue in the workplace. As demonstrated, Eqn 10 triggers the alarm when appropriate for all incidents of fatigue where the participant scores over 6 on the KSS (P1, P3, P8, P15). Of note, the alarm does not trigger for P1 or P8 due to the restorative effects of break administration. The rows highlighted in orange are >1 KSS score away from the predicted KSS, the rows highlighted in green are <1 KSS score away from the predicted KSS. Participants 1 and 13 have been examined recently as to their false alarms. Participant 6 was an example of an individual who worked overtime, as such breaching the max parameter for the time on shift variable.

Table 7.3: A comparison of the reported KSS scores when the alarm would have triggered on shift for Eqn 9 and Eqn 10.

Participant	Eqn 9	Eqn 10	Max KSS of participant
1	N/A	4.24	6
2	N/A	N/A	4
3	5.44	5.60	7
6	4.55	4.55	5
8	N/A	6.41	8
9	N/A	N/A	4
10	N/A	N/A	4
11	N/A	N/A	3
12	N/A	N/A	4
13	3.33	3.65	4
14	N/A	N/A	5
15	5.08	5.37	8

Below (Figure 7.7) is an example where the alarm is not required due to low RMSSD throughout the participants shift. In most of the examples listed in Table 7.3, this is the reoccurring case, when paired with good sleep quality. All day-shift participants in this research did not score higher than a 5, therefore it is important that the algorithm isn't triggering alarms when not necessary.

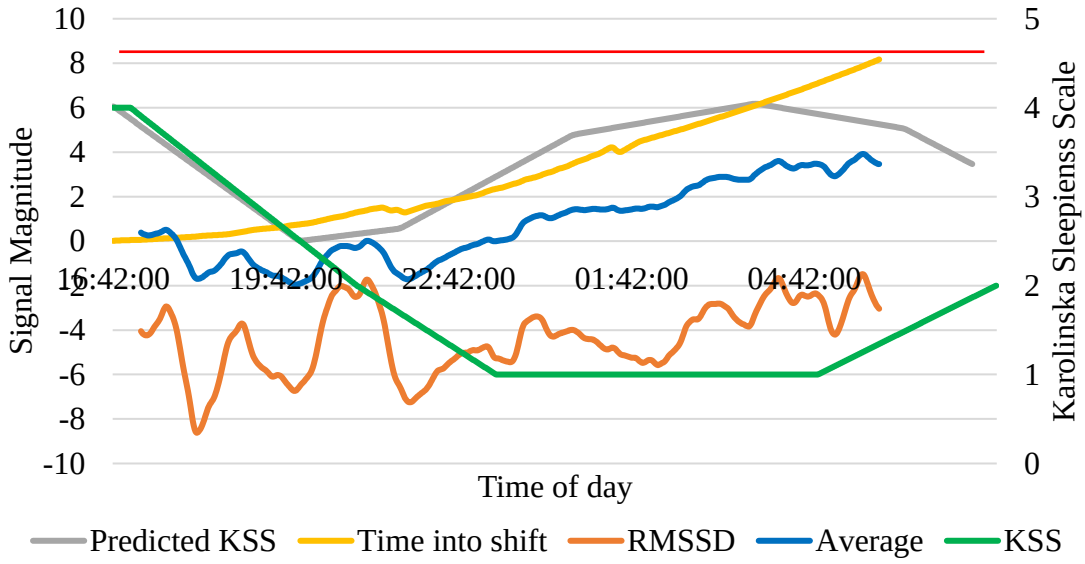


Figure 7.7: The alarm not being required for participant 2, due to good sleep score and low RMSSD levels throughout the shift, using Eqn 10.

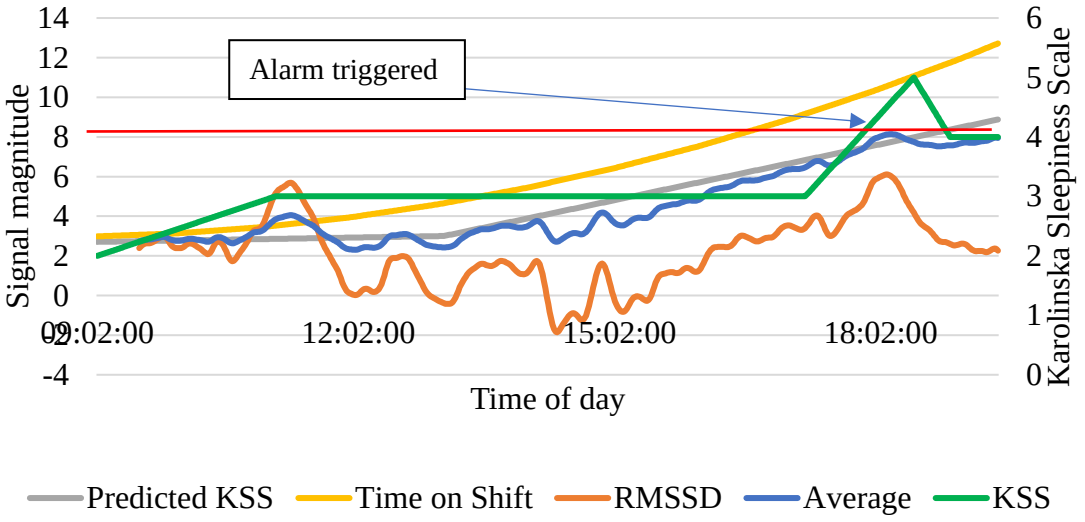


Figure 7.8: The alarm being triggered for participant 6, due to average sleep score and working overtime, despite low RMSSD levels throughout the shift, using Eqn 10.

Figure 7.8 reflects the alarm trigger (blue circle) for participant 6, as the participant is working more than 2 hours after the end of their scheduled shift time. Of note, the RMSSD measure increases towards the end of the shift, which is the catalyst for the alarm trigger in this instance.

7.4.1 Validating the algorithms

When comparing the Karolinska Sleepiness Scale (KSS) to the algorithm, Bland-Altman plots, mean squared error (MSE), and mean absolute error (MAE) are commonly used metrics for these comparisons.

A Bland-Altman plot (Giavarina, 2015) is a graphical representation of the agreement between two methods of measurement, in this case, the KSS and the algorithm's predictions. It displays the difference between the two measurements (y-axis) against their mean (x-axis), allowing for the visualization of any systematic bias or variation in the differences between the two methods. In this instance, the difference between the algorithm and KSS is measured. The plot can reveal if the differences are consistent across the range of measurements and if there are any outliers or trends in the data. The Bland-Altman plot is especially useful when comparing the performance of two different measurement methods or instruments, as it provides a way to assess the agreement between them beyond simple correlation or regression analyses. It can help identify any systematic differences between the two methods and assess the level of agreement between them.

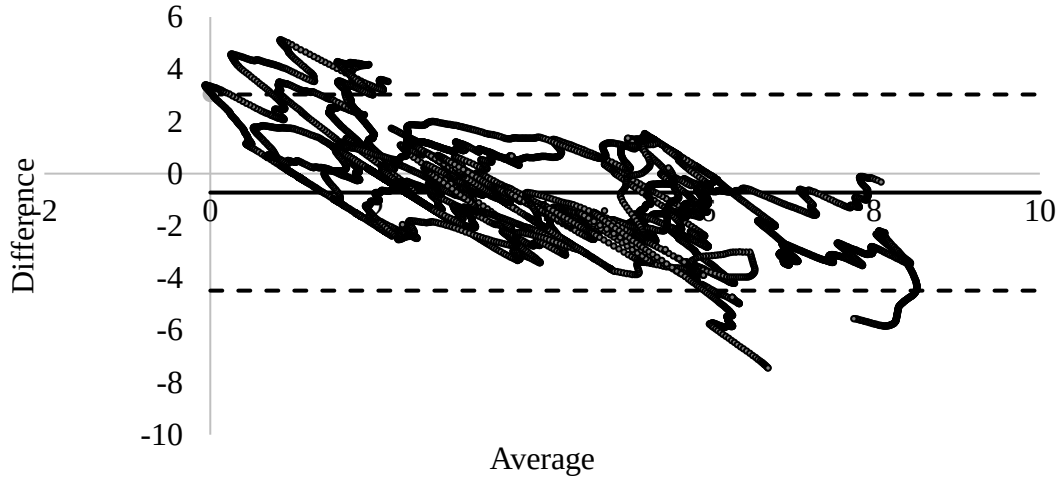


Figure 7.9: Bland-Altman plot comparing KSS and *L* within 95% confidence intervals for Eqn 9.

within ±1.96 Standard Deviations of the mean – limits of agreement. The Bland-Altman plot for Eqn 10 (Figure 7.10) has 91.10% of the results fall within ±1.96 Standard Deviations of the mean – limits of agreement. This measure accurately predicts the level of KSS of surgeons on shift, mitigating the need for filling out questionnaires on shift. This method can be beneficial when surgeons may not truthfully report their fatigue levels on shift, or fight through the effects of fatigue without considering a break.

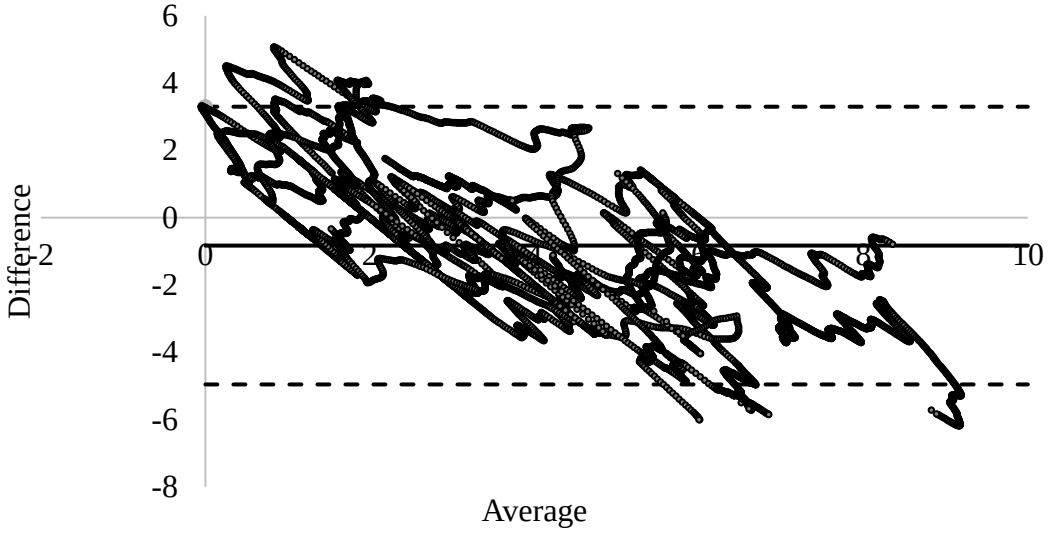


Figure 7.10: Bland-Altman plot comparing KSS and *L* within 95% confidence intervals for Eqn 10.

MSE is a measure of the average squared difference between the predicted values and the true values. It provides a quantitative estimate of the accuracy of the algorithm's predictions and is calculated as the mean of the squared differences between the predicted and true values. MSE is sensitive to outliers and can be useful for comparing the performance of different algorithms. For Eqn 9 and 10, the MSE values are 0.1135 and 0.11785, respectively. MAE is a measure of the average absolute difference between the predicted values and the true values. Like MSE, it provides a quantitative estimate of the accuracy of the algorithm's predictions, but it is less sensitive to outliers. MAE is calculated as the mean of the absolute differences between the predicted and true values. For Eqn 9 and 10, the MAE values are 0.11228 and 0.11431, respectively.

The values for MSE and MAE are relatively similar between the two equations, demonstrating that the algorithms predictions of the KSS are off by approximately 0.11, demonstrating considerable accuracy between the two measures.

If 88-91% of the algorithm's predictions fall within +- 1.96 standard deviations of the KSS scores, it is a good level of agreement between the algorithm's predictions and the true KSS scores. In statistical terms, +- 1.96 standard deviations from the mean correspond to the 95% confidence interval for a normally distributed variable. This means that if the algorithm's predictions fall within this range for 88-91% of the cases, there is evidence that the algorithm is providing accurate predictions. Different fields and applications may have different standards for what is considered acceptable levels of agreement between an algorithm and the true values. Additionally, other evaluation metrics such as mean squared error or mean absolute error may provide additional information about the accuracy of the algorithm's predictions.

7.5 Application of advisory system

As aforementioned, the advisory system signals an alarm when the *L* value reaches a value of 8.14 (Eqn 9) or 8.15 (Eqn 10). However, some participants exceed this threshold again after the administration of a break later in their shift, particularly 24-hour shifts. For example, administering the first break after reaching the alarm threshold may signal the next alarm to trigger when the *L* reaches 9.69 (Eqn 9) or 9.81 (Eqn 10), the equivalent of seven on the KSS (Table 6.2). As such, further recommendations for the application of the advisory system may result in linearly repeated alarms, such as the *L* equivalent of the KSS scale in Table 6.2. These alarm triggers are indicated in Figure 7.11 with red circles.

Figure 7.11 shows Participant 3 with the multiple alarm triggers with reference to Eqn 10. The first alarm would trigger when the *L* value reaches 8.15. In this instance, the alarm is triggered 16:08pm, 500 minutes into the shift. The following alarm will sound at 9.81, at 00:00am, 972 minutes into the shift. In this case, 473 minutes after the participant’s last alarm, the alarm triggers again due to the increase in predicted KSS and Time on Shift parameters. In this example, the participant has minimal restorative

rest periods. A final alarm would trigger at 11.47, 02:44am, 1136 minutes into the shift, when the KSS is predicted to be 8.

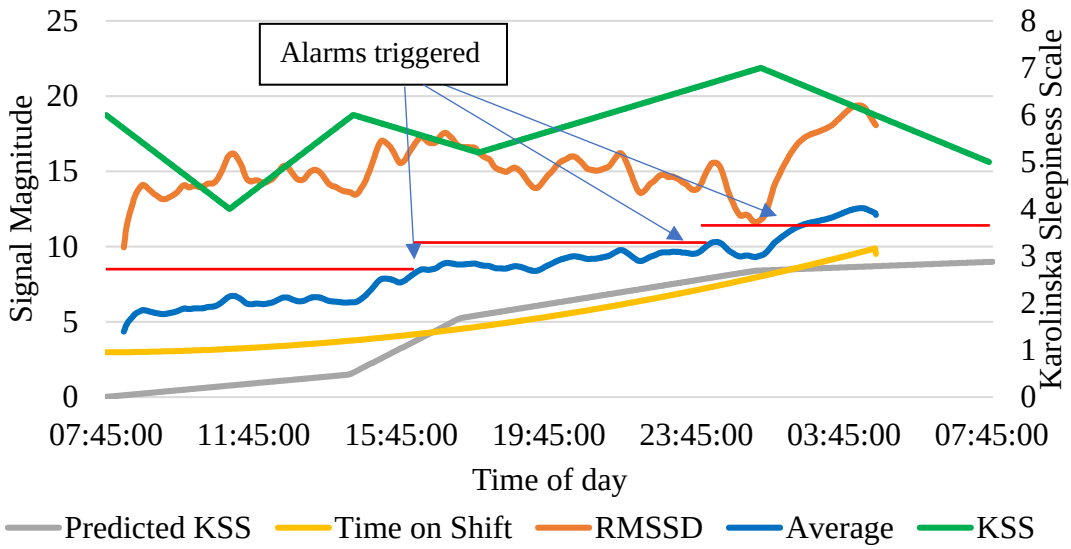


Figure 7.11: Signal magnitude showing example of participant with multiple alarms triggering fatigue prompts due to high RMSSD levels (Figure 5.9) using three different parameters (Time on shift, RMSSD, predicted KSS), compared against Karolinska sleepiness scale.

The application of multiple alarms at increasing *L* magnitudes is due to the continual rise in the Time on Shift variable, as the average *L* parameter will always increase. The participant had an average sleep in this scenario, offsetting the time on shift variable higher up the y-axis. The RMSSD of this participant is relatively high throughout the shift, as well as the actual KSS. In this example, the participant's lowest KSS score is above 5 when the alarm sounds (neither alert nor sleepy), therefore they are at risk of making errors in the workplace.

Figure 7.11 can be used to visualise the alarm system as the *L* value subsequently increases in Table 6.2. Figure 7.12 shows that if a break alarm is silenced (or not taken deliberately), the next subsequent alarm threshold is prioritised. This may be used when the recipient cannot take a structured break soon or near the end of their shift. However, if the alarm is not silenced, then the alarm may continue to propose an advisory break to the recipient. This may be used when the recipient is close to completing a task.

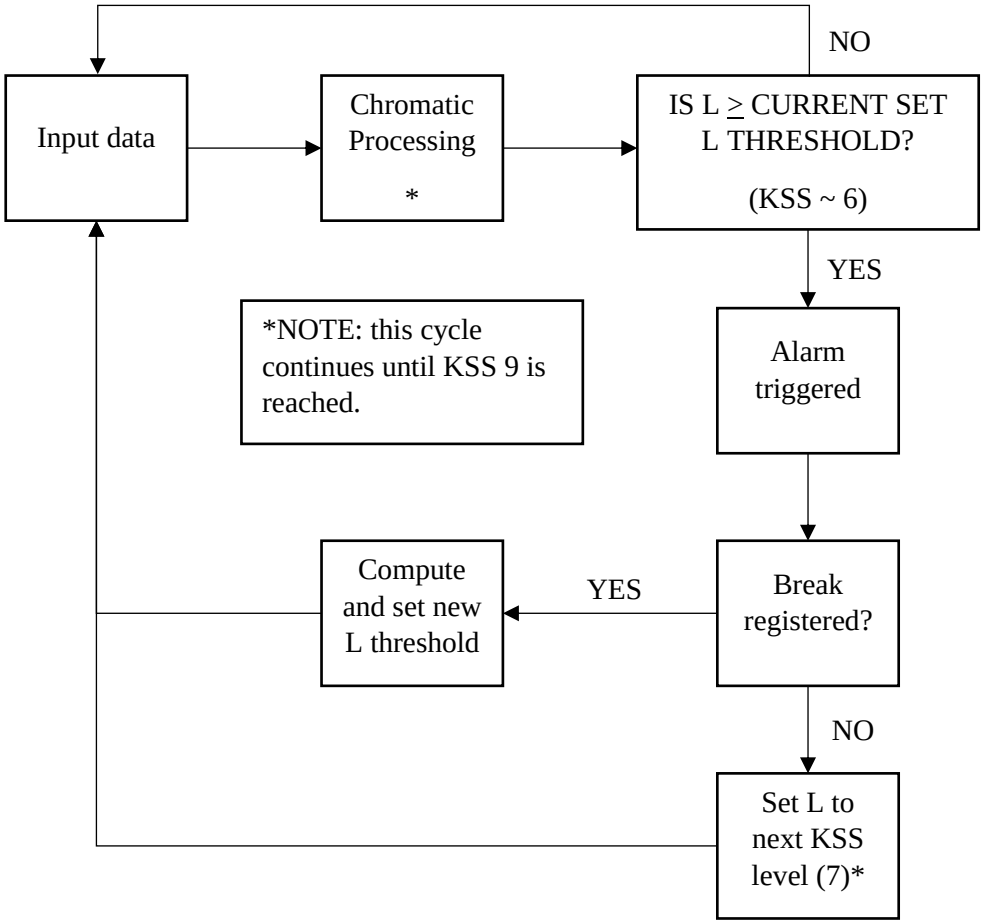


Figure 7.12: Flowchart for the administration of multiple alarms for advisory system using Table 6.2 for alarm thresholds, with input data based on Eqn 9 and 10 in 1-minute epochs.

Figure 7.13 displays a flowchart for what happens in the chromatic processing block in order to compute the resultant L threshold (outlined with an Asterix in the box). It shows two branches from the data sources depending on the equation used, which produces the output that carries on the flow chart in Figure 7.13.

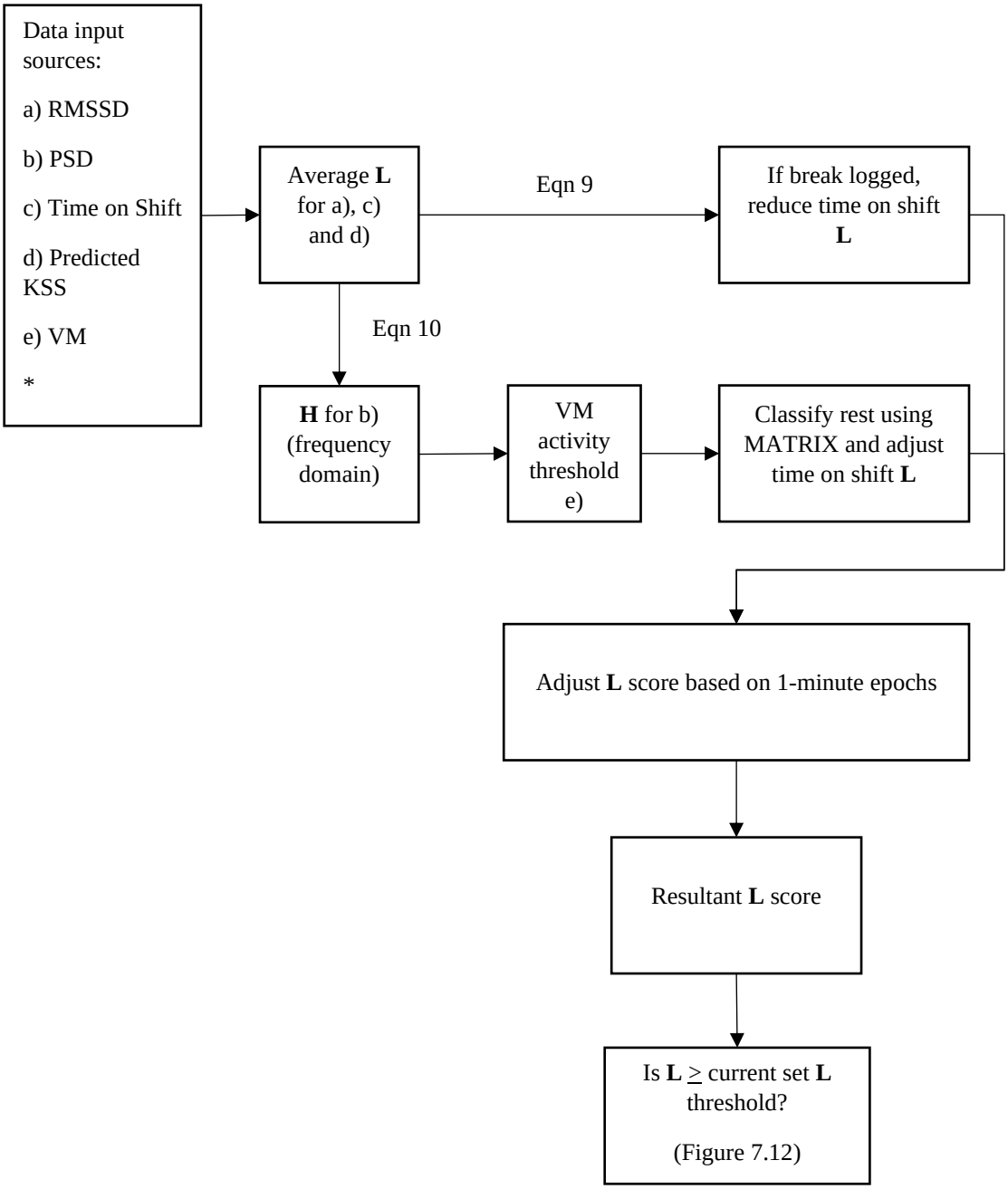


Figure 7.13: Flowchart which shows in depth analysis of the chromatic processing block displayed in Figure 7.12.

7.6 Summary

This section applies the algorithm into practice, advising the user to take a break when necessary. Similar trends can be identified between RMSSD and KSS scores, further validating the relationship between the two, though more data must be collected to strengthen these findings.

Using a chromatic approach, an algorithm has been developed which identifies the signal magnitude of RMSSD, predicted KSS and time on shift. The average result is used and when the average exceeds the alarm threshold, it is triggered. This is the first algorithm of it's kind to use physiological parameters to prompt the user to take a break on shift, with the intention of reducing errors made in the workplace and therefore increasing patient outcomes. This chapter compared two algorithms, and found that the second algorithm (eqn 10) was better at identifying fatigue with appropriate alarm prompts and also a better correlation between the algorithm and KSS.

However there are limitations to the algorithm which need to be addressed with future research. Firstly, a sleep tracker is required so that the quality of the sleep can be objectively measured, preventing false records or ambiguous responses. Of course, KSS is not the most accurate measure of sleepiness, however without being able to use state-of-the-art technology such as electroencephalography (EEG) due to it's invasiveness, an alternative solution is used. The subjectiveness of the KSS scale ultimately results in discrepancies amongst individuals as to how they are feeling at any given moment. It is difficult to scale fatigue, as it is sleepiness or pain. As such, interpersonal differences cause large variation in the KSS. Self-perceived sleepiness can vary more considerably than every three hours on shift. However in this work, three hours was chosen so that it does not become too invasive for the surgeons. It is presumed in this research that the KSS is linear, when this may not be the case.

Two algorithms (Eqn 9 and 10) shows promise at identifying the moment where surgeons are fatigued (or not) on shift. One requires user input (Eqn 9), however the other one (Eqn 10) utilises novel application of chromatics for characteristics of the PSD of HRV to determine an individuals autonomic nervous system activity. Both algorithms are very reliable at predicting the KSS, with a MAE of 0.11 for both. However without real-time deployment into the field, the algorithm cannot be accurately tested for its validity.

When the alarm triggers, taking the opportunity to relax is highly recommended. This relaxation can take the form of meditation or simply stepping away from one's desk for a few minutes. Relaxation techniques such as meditation or mindfulness can help to reduce stress and promote a sense of calmness and well-being, which can in turn improve productivity and performance. Stepping away from one's desk and taking a short break can also help to alleviate mental and physical fatigue and reduce the risk of burnout. Therefore, taking advantage of the opportunity to relax when the alarm triggers can be an effective strategy for maintaining a healthy work-life balance and promoting overall well-being.

Chapter 8

Conclusions and future work

8.1 Conclusions

The main objective of this research was to develop a real-time holistic and inexpensive fatigue monitoring system, which can be utilised in the workplace. The advisory system should be able to detect early signs of fatigue, advising the user to take a break when possible to alleviate the possibility of errors that are common in shift work. The advisory system is based on physiological indicators from a pilot study (section 3.3) and further experimental results (section 5.4), with the application of novel chromatic processing techniques to extract the required information for the prediction of surgeon fatigue with further visual representation in the form of chromatic polar plots, particularly for the frequency-domain heart rate variability (HRV) power spectral density (PSD) analysis.

The physiological variables portrayed in the algorithm include Heart Rate Variability (HRV) with applications of both time and frequency domain, and circadian rhythm. In addition, the Time on Shift and sleep quality make up the final variable in the algorithm. The research is separated into two algorithms, one that requires user input to denote when breaks are being taken and one which uses the chromatic parameter *Hue* to differentiate between sympathetic and parasympathetic nervous activity when looking at the PSD of HRV in unison with vector magnitude (VM). In both algorithms, they

accumulate the average *strength* of these scores to provide the chromatic output L. The variables in the algorithm are individually weighted, to be indicative of the Karolinska Sleepiness Scale (KSS), with high statistical correlation. In the first instance, the algorithm alerts the user when they are predicted a score of '6' ("some signs of sleepiness") so that a break may be taken to reduce the chances of the participant scoring a '7' ("moderately sleepy"). In the circumstance that the participant takes a break or rejects the alarm, it may sound again when the participant's predicted a '7', and so forth.

With the use of chromatic analysis for both algorithm, data from the algorithm can be presented using polar plots. The chromatic parameters *H* and *L* represent the category of tiredness and the tiredness level, respectively. A *H* value closer to 60° represents a low tiredness level and minimal risk of fatigue, whereas a *H* closer to 300° represents greater risk of fatigue (Figure 6.7). Below 180° indicates the decline and possible peak of at least one of the three variables in the algorithm and therefore is a safe zone for the user to be in. Data points over 240° indicate there is at least two increasing variables and one possible peak of the variables and as such puts the user at an increased risk of fatigue. Coupled with the *L* parameter, chromatic analysis indicates the magnitude of the users fatigue risk.

For both algorithms (Eqn 9 and 10), 88.06% and 91.10% of the results fall within $\pm$1.96 Standard Deviations of the mean – limits of agreement of the actual KSS scores, indicating that the algorithms can accurately predicts the level of KSS of surgeons on shift, mitigating the need for filling out questionnaires on shift. Eqn 10 mitigates the user input of taking breaks on shift and appropriately monitors 'true' restorative events, as outlined by the chromatic *H* parameter of the PSD and the VM. The values for mean squared error and mean absolute error are relatively similar between the two equations, demonstrating that the algorithms predictions of the KSS are off by approximately 0.11, providing considerable accuracy between the two measures. Both equations work exceptionally at predicting workplace fatigue, however Eqn 10 does not require user input, therefore mitigates the possibility of user error when falsely reporting break opportunities. However, it is important for working establishments to have environments where shift workers are able to benefit

from true restorative events, as they are seldom during periods where participants said they were sleeping.

The advisory system may be used in circumstances where surgeons (or other shift workers) have worked for a prolonged period without a break. Similarly, surgeons may be at a heightened risk of fatigue due to the time of day or the relaxed nature observed from the RMSSD and it is advised that a break be taken to further mitigate the chances of errors occurring in the workplace. The system is advisory, thus it would not act as a mandatory requirement of the individual to cease their current activities in the workplace. However, in instances where the alarm is ignored it may trigger again at the subsequent iteration presented in Table 6.2, as demonstrated in Figure 7.12. There are instances where surgeons may not be able to take a break for prolonged periods due to an operation of considerable length, in which case the surgeon may be recommended to take a break following completion of the operation.

There are ethical considerations with this research. As mentioned in Table 3.4, not all individuals would be willing to wear the vest (or chest strap sensor) and it may appeal to certain demographics more than others. As such, the administration of such a device would not be made compulsory, but offer opportunity to those who relate to the 'concept of dynamic scheduling'. This is where workers take breaks at the most beneficial opportunity in the fatigue cycle, rather than scheduled by the company in rotation. In most instances however, scheduled breaks may be too late or too early and lack beneficial effects. The advisory system offers the user the opportunity to make them aware of their alertness, using a proactive approach to mitigate fatigue onset further.

The present research identifies surgeons at their most fatigued when their RMSSD is high. However, with reference to Eqn 10, RMSSD (time-domain) and the PSD (frequency-domain) capture different aspects of heart rate variability (HRV), where RMSSD primarily reflects parasympathetic activity, while the R dominant filter (indicative of a high LF/HF ratio) assesses the relative balance between sympathetic and parasympathetic activity. While parasympathetic activity is typically associated

with higher RMSSD values, sympathetic activity can also contribute to LF/HF ratio. It may be beneficial for the individual, upon triggering an alarm, to take opportunity to enter a different headspace, escape their current environment and have opportunity to consume a caffeinated beverage or nap for at least 5 minutes for the restorative effects to be beneficial.

There were three main aims of this thesis. The first aim, to identify contributors to fatigue within the workplace and possible mitigation strategies was completed with a questionnaire administered to Surgeons, which found that the most common contributors were high levels of demand, persistent mental activity and prolonged static postures. The main mitigators were taking a break where possible, changing positions at work/performing surgery and adjusting the surgical field. The questionnaire also highlighted gender differences with workplace fatigue. The second aim was to produce an advisory system to predict the onset of fatigue. This was achieved using chromatics and several inputs to quantify a predicted fatigue score against the KSS. The final aim, utilising chromatic processing, was achieved using the second aim with great success.

This thesis concludes that it is possible to design a fatigue-monitoring system based on physiological parameters for detecting the early fatigue signs in the workplace, with only a chest strap, so that it is unobtrusive to the user in field-based work. The novelty of this thesis is provided with pioneering technology, which can monitor fatigue scores of human in the workplace, in real time. This has not yet been achieved but requires larger testing to solidify this methodology.

8.2 Future work

Reflective direction for future work would include longitudinal studies which may be able to provide much richer data over multiple shifts, for a more comprehensive understanding of the fatigue phenomena and its evolution over time. Additionally machine learning techniques may be an alternative solution to gain deeper insights into complex data sets.

The experiments were conducted before the construction of the advisory system and therefore the algorithm has not yet been tested in a real time setting. As such, the development of a mobile application including the algorithm and Bluetooth capability can be implemented in the workplace, when paired with a chest strap, for the monitoring of real-time surgeon fatigue. Further development of the algorithm could introduce machine learning that can more accurately predict the onset of fatigue, through user inputs of the KSS on the application, modifying the parameters for HRV normalisation accordingly.

The sample size in this study is small and future research should recruit a larger, more varied cohort from multiple hospitals. Participants voluntarily took part in the study and therefore surgeons who are currently experiencing fatigue may have been more likely to participate in the study.

The chromatic parameters S and L may yield important information as to its association with frequency-domain measures, such as the LF/HF ratio and must be investigated further, to provide a more robust algorithm with the possibility to reduce computational cost.

The cohort chosen in this thesis were surgeons, however the effects of the algorithm can be utilised across all shift-workers. Yet, the predicted KSS variable introduced into the algorithm is specific to the general trends of sleepiness experienced by surgeons and as such should be re-evaluated for different cohorts. Similarly, the surgeons did not work within 24-hours of their shifts in this study. Future research may consider the longitudinal effects of fatigue for shift workers who work multiple days in a row, for example identifying circadian rhythm trends across multiple days and analysing sleep patterns in between shifts.

The thesis highlights the benefits that decreased time-domain HRV has on reducing predicted KSS. However, there are variables that require further investigation with the

development of the proposed technology that may directly affect HRV. These variables include smoking (Sondermeijer et al., 2002), caffeine (Rauh et al., 2005) or other prescribed medications. The effects of smoking are not included in this as no participants were smokers. Similarly, mental health and menstrual cycles in females also require further investigation as they can effect both feelings of fatigue and HRV.

Of note, circadian rhythms and sleep quality can vary considerably between individuals. Whilst Eqn 10 mitigates the user input for break administration, sleep quality has three options, 'Good', 'Average' and 'Poor', which are not objective measures that yield more accurate representation of the users sleep quality. In order to further improve the algorithm, non-intrusive sleep technology such as Actigraphy or sleep tracking mats with Ballistocardiography (Surantha et al., 2020) can be utilised for a more detailed examination of the sleep quality so that it does not remain subjective. Alternatively, a Likert-scale with more than three options (e.g. five) may be used in situations where users are unsure on the quality of their sleep, to improve this estimation parameter further. Then the algorithm will be free of user input, and able to be deployed into health-care (or other shift work) professions to reduce the errors made in the workplace.

References

1. Actiheart. *Actiheart 5*. Available at https://www.camntech.com/actiheart-5/?gclid=Cj0KCQjws4aKBhDPARIsAIWH0JWqejayPxgzPmNjPHiJAZnqrYaWTD5KeHg8Ac5WPe9eaUO_l_c9JAUaAgDzEALw_wcB Accessibility verified September 15 2021.
2. Aeschbacher, S., Schoen, T., Dorig, L., Kreuzmann, R., Neuhauser, C., Schmidt-Truksass, A., Probst-Hensch, N. M., Risch, M., Risch, L. and Conen, D. (2017) ,Heart rate, heart rate variability and inflammatory biomarkers among young and healthy adults.' *Annals of medicine*, 49(1), pp. 32-41.
3. Alleblas, C. C. J., Marie de Man, A., van den Haak, L., Vierhout, M. E., Jansen F, W. and Nieboer, T. E. (2017) 'Prevalence of Musculoskeletal Disorders Among Surgeons performing Minimally Invasive Surgery: A Systematic Review.' *Annals of Surgery*, 226(6), pp. 905-920.
4. Amirian, I., Danielsen, A. K. and Rosenberg, J. (2013) 'Perception of fatigue among surgeons during night shifts.' *Annals of the royal college of surgeons of England*, 95(7), pp. 1-5.
5. Anwer, S., Heng, L., Antwi-Afari, M. F., Umer, W. and Wong, A.Y. L. (2021) 'Evaluation of physiological metrics as a real-time measurement of physical fatigue in construction workers: state-of-the-art review.' *Journal of construction engineering and management*, 147(5); https://doi.org/10.1061/(ASCE)CO.1943-7862.0002038
6. Arendt, J. (2010) 'Shift work: coping with the biological clock.' *Occupational medicine*, 60(1), pp. 10-20.
7. Arora, V., Dunphy, C., Chang, V. Y., Ahmad, F. Humphrey, H. J. and Meltzer, D. (2006) 'The effects of on-duty napping on intern sleep time and fatigue.' *Annals of internal medicine*, 144, pp.792-798.
8. Aryal, A., Ghahramani, A. and Becerik-Gerber. B. (2017) 'Monitoring fatigue in construction workers using physiological measurements.' *Autom. Constr*, 82, pp. 154-165.

9. Babkoff, H., Caspy, T., Mikulincer, M. and Sing, H. (1991) 'Monotonic and rhythmic influences: a challenge for sleep deprivation research. *Psychology Bulletin,* 109 pp. 411-428.
10. Bae, S. H., Trinkoff, A., Jing, H. and Brewer, C. (2013) 'Factors associated with hospital staff nurses working on-call hours: a pilot study' *Workplace Health and Safety,* 61(5), pp. 203-211.
11. Baek, H. J., Cho, C. H., Cho, J. and Woo, J.M. (2015) 'Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability.' *Telemedicine and E-Health,* 21, pp. 404–414.
12. Ban, K. A., Chung, J. W., Matulewicz, R. S., Kelz, R. R., Shea, J. A., Dahlke, A. R., Quinn, C. M., Yang, A. D. and Bilimoria, K. Y. (2021) 'Gender-Based Differences in Surgical Residents' Perception of Patient Safety, Continuity of Care, and Well-Being: An Analysis from the Flexibility in Duty Hour Requirements for Surgical Trainees (FIRST) Trial.' *American College of Surgeons,* 12 pp. 126-139.
13. Basner. M., Mollicone, D., Dinges, D. F. (2011) 'Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation.' *Acta Astronautica,* 69, pp.949-959.
14. Berry, R. B. and Wagner, M. H. (2015) 'Introduction' *Sleep medicine pearls (third edition)* pp. 181-187.
15. Berwick, D. M. and Leape, L. L. (1999) 'Reducing errors in medicine.' *British medical journal*, 319(136), pp. 136-137.
16. Bethune, R. and Francis, N. (2015) 'The human factor in minimal access surgical taining: how conscientious, well trained surgeons make mistakes.' *Training in minimal access surgery*, pp. 197-208.
17. Boksem, M. A. S. and Tops, M.(2008) 'Mental fatigue: costs and benefits.' *Brain research reviews,* 59(1) pp. 125-139.
18. Boksem, M. A. S., Meijman, T. F. and Lorist, M. M. (2005) 'Effects of mental fatigue on attention: An ERP study. *Cognitive brain research,* 25, pp. 107-116.
19. Borragan, G., Slama, H., Bartolomei, M. and Peigneux, P. (2017) 'Cognitive fatigue: A time-based resource-sharing account.' *Cortex,* 89, pp. 71-84.

20. Bosch, C. and Sonnentag, S. (2019) ‘Should I take a break? A daily reconstruction study on predicting micro-breaks at work.’ *International journal of stress management*, 26(4) pp. 378-388.
21. Bowman, D. S., Hanowski, R. J., Alden. A., Gupta, S., Wiegand, D., Baker, S., Stanley, L. M. and Wierwille, W. (2012) ‘Development and assessment of a driver drowsiness monitoring system.’ *Federal motor carrier safety administration*, pp. 1-207.
22. Caruso, C. C., Baldwin, C. M., Berger, A., Chasens, E. R., Landis, C., Redeker, N. S., Scott, L. D. and Trinkoff, A. (2017) ‘Position statement: Reducing fatigue associated with sleep deficiency and work hours in nurses.’ *Nursing Outlook*, 65 pp. 766–768.
23. CNBC (2018) *The third-leading cause of death in US most doctors don’t want you to know about.* Available at: https://www.cnbc.com/2018/02/22/medical-errors-third-leading-cause-of-death-in-america.html#:~:text=A%20recent%20Johns%20Hopkins%20study,after%20heart%20disease%20and%20cancer. (Accessed: 03/12/2022)
24. Cockerham, M., Kang, D.H., Howe, R., Welmer, S., Boss, L. and Kamat, S. R. (2018) ‘Stress and cortisol as predictors of fatigue in medical/surgical nurses and nurse leaders: A biobehavioural approach’ *Journal of nursing education and practice*, 8(5), pp. 76-83.
25. Dahlke, A. R., Johnson, J. K., Greenberg, C. C., Love, R., Lindsey, K., Hewitt, D. B., Quinn, C. M., Engelhardt, K. E. and Bilimoria, K. Y. (2018) ‘Gender Differences in Utilization of Duty-Hour Regulations, Aspects of Burnout, and Psychological Well-being Among General Surgery Residents in the United States.’ *Annals of Surgery*, 268 pp. 204-211.
26. Dawson, D., Darwent, D. and Roach, G. R. (2017) ‘How should a bio-mathematical model be used within a fatigue risk management system to determine whether or not a working time arrangement is safe?’ *Accident: analysis and prevention*, 99, pp. 469-473.
27. Dawson, D., Noy, I. and Harma, M. L. (2011) ‘Modelling fatigue and the use of fatigue models in work settings.’ *Accident: analysis and prevention*, 42(2), pp. 549-564.

28. Dawson, D., Searle, A. K. and Paterson, J. L. (2014) ‘Look before you (s)leep: evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry.’ *Sleep medicine reviews*, 18(2), pp. 141-152.
29. Deakin, A. G., Rallis, I., Zhang, J., Spencer, J. W. and Jones, G. R. (2006) ‘Towards Holistic Chromatic Intelligent Monitoring of Complex Systems.’ *Sensor Review*, 26(1), pp. 11-17.
30. Dias, R. D., Conboy, H. M., Gabany, J. M., Clarke, L. A., Osterweil, L. J., Arney, D. and Zenati, M. A. (2018). Intelligent interruption management system to enhance safety and performance in complex surgical and robotic procedures. In *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis* (pp. 62-68). Springer, Cham.
31. Dorion, D. and Darveau, S. (2013) ‘Do micropauses prevent surgeon’s fatigue and loss of accuracy associated with prolonged surgery?’ *Annals of surgery*, 257, pp. 256-259.
32. Engelmann, C., Schneider, M., Kirschbaum, C., Grote, G., Dingemann, J., Schoof, S. and Ure, B. M. (2011) ‘Effects of intraoperative breaks on mental and somatic operator fatigue: a randomised clinical trial.’ *Surgical Endoscopy*, 25, pp. 1245-1250.
33. Farquharson, B., Allan, J., Johnston, D., Johnston, M. Choudhary, C. and Jones, M. (2012) ‘Stress amongst nurses working in healthcare telephone-advice service: relationship with job satisfaction, intention to leave, sickness absence, and performance.’ *Journal of advanced nursing*, 68, pp. 1624-1635.
34. Freedman, D., Pisani, R., Purves, R. 2007. Statistics (international student edition). *Pisani, R. Purves, 4th edn.* WW Norton & Company, New York. pp. 119-157.
35. Gates, M., Wingert, A., Featherstone, R., Samuels, C., Simon, C. and Dyson, M. P. (2018) ‘Impact of fatigue and insufficient sleep on physician and patient outcomes: a systematic review.’ *British medical Journal Open*, 8(9) e021967.
36. Giavarina, D. (2015) ‘Understanding Bland Altman analysis.’ *Biochemical medicine*, 25(2) pp. 141-151.
37. Gonzalez, L., Sasangohar, F., Mehta, R. K., Lawley, M. and Erraguntla, M. (2017) ‘Measuring fatigue through heart rate variability and activity

recognition: a scoping literature review of machine learning techniques. *Proceedings of the human factors and ergonomics society annual meeting,* 61(1), pp. 1748-1752.

38. Graversen, J. A., Korets, R., Mues, A. C., Katsumi, H. K., Badani, K. K., Landman, J. and Gupta, M. (2011) 'Prospective randomised evaluation of gel mat foot pads in the endoscopic suite.' *Journal of endourology*, 25(11) pp.1793-1796.
39. GreenTEG. *Core*. Available at https://shop.greenteg.com/core Accessibility verified September 15 2021.
40. Griffiths, C., Bowen, J. and Hinze, A. (2017) 'Investigating Wearable Technology for Fatigue Identification in the Workplace.' *Human-Computer Interaction - INTERACT 2017*, pp. 370-380.
41. Gui, J. L., Nemergut, E. C. and Forkin, K. T. (2021) 'Distraction in the operating room: A narrative review of environmental and self-initiated distractions and their effect on anaesthesia providers.' *Journal of clinical anesthesia,* 68:, doi.org/10.1016/j.jclinane.2020.110110.
42. Haghighi, S. K. and Yazdi, Z. (2015) 'Fatigue management in the workplace.' *Independent Psychiatry Journal,* 24, pp. 12-17.
43. Hallbeck, M. S., Lowndes, B. R., Bingener, J., Abdelrahman, A. M. Yu, D., Bartley, A. and Park, A. E. (2017) 'The impact of intraoperative microbreaks with exercises on surgeons: A multi-center cohort study.' *Applied ergonomics,* 60 pp. 334-341.
44. Haramis, G., Rosales, J. C., Palacios, J. M., Okhunov, Z., Mues, A. C., Lee, D., Badani, K., Gupta, M. and Landman, J. (2010) 'Prospective randomised evaluation of foot gel pads for operating room staff comfort during laparoscopic renal surgery.' *Laparoscopy and robotics,* 76(6) pp. 1405-1408.
45. Harris, J. D., Staheli, G., LeClere, L., Andersone, D. and McCormick, F. (2015) 'What effects have resident work-hour changes had on education, quality of life anf dafety? A systematic review.' *Clinical orthopaedics and related research.* 473 pp. 1600-1608.
46. Hevezi, J. A. (2015) 'Evaluation of a meditation intervention to reduce the effects of stressors associated with compassion fatigue among nurses.' *Journal of Holistic Nursing*, 3, pp. 343-350.

47. Hinz, M., Stein, A. and Uncini, T. (2011) ‘validity of urinary monoamine assay sales under the “spot baseline urinary neurotransmitter testing marketing model”’ *International journal of nephrology and renovascular disease,* 4, pp. 101-113.
48. Hirshkowitz, M. and Sharafkhaneh, A. (2017) ‘Chapter 169 – Evaluating Sleepiness’ *Principles and practice of sleep medicine (sixth edition)*, pp. 1651-1658.
49. Hockey, G. R. J. (1997) Compensatory control in the regulation of human performance under stress and high workload: a cognitive energetical framework. *Biological psychology,* 45, pp. 73-93.
50. Hockey, G. R. J. (2011) ‘A motivational control theory of cognitive fatigue. In: Cognitive fatigue: Multidisciplinary perspectives on current research and future applications.’ *APA,* pp.167–187; https://doi.org/10.1037/12343-008
51. Hockey, R. J. G. and Earle, F. (2006). ‘Control over the scheduling of simulated office work reduces the impact of workload on mental fatigue and task performance.’ *Journal of experimental psychology: applied,* 12(1), pp. 50-65.
52. Holding, D. (1983) Stress and Fatigue in Human Performance, John Wiley and Sons, Durnham, pp. 145-164.
53. Holdsworth, L. and Evens, T. (2017) ‘Is heart rate variability a useful marker of stress and fatigue in emergency and pre-hospital clinicians? - a systematic review.’ *Resuscitation,* 118(1), pp. 95.
54. Horne, J. A. 1988. *Why we sleep: the functions of sleep in humans and other mammals.* Oxford University Press, Oxford.
55. Horne, J. A. and Reyner, L. A. (1996) ‘counteracting driver sleepiness: effects of napping, caffeine and placebo.’ *Psychophysiology*, 33(3), pp. 306-309.
56. Houdmont, J., Daliya, P., Theophilidou, E., Adiamah, A., Hassard, J. and Lobo, D. N. (2021) ‘Burnout among surgeons in the UK during the covid-19 pandemic: A cohort study.’ *Original scientific report*, 49 pp. 1-9.
57. Hulsheger, U. R. (2016) ‘From dawn till dusk: shedding light on the recovery process by investigating daily change patterns in fatigue.’ *Journal of applied psychology*, 101(6), pp. 905-914.
58. IBM Corp. (2020). IBM SPSS Statistics for Windows (Version 27.0) [Computer software]. IBM Corp.

59. James, F. O., Waggoner, L. B., Weiss, P. M., Patterson, D., Higgins, S., Lang, E. S. and Van Dongen, P. A. (2018) 'Does Implementation of Biomathematical Models Mitigate Fatigue and Fatigue-related Risks in Emergency Medical Services Operations? A Systematic Review.' *Prehospital Emergency Care*, 22, pp. 69-80.
60. Jarrad, R., Hammad, S., Shawashi, T. and Mahmoud, N. (2018) 'Compassion fatigue and substance use among nurses.' *Annals of general* psychiatry, 17(13), https://doi.org/10.1186/s12991-018-0183-5
61. Jasper, I., Haubler, A., Marquardt, C. and Hermsdorfer, J. (2009) 'Circadian rhythm in handwriting.' *Journal of sleep research,* 18, pp. 264-271.
62. Johns, M. W. (2010) 'A new perspective on sleepiness.' *Sleep and biological rhythms*, 8, pp. 170-179.
63. Johnston, D.W., Allan, J.L. and Powell, D.J.H et al. (2018) 'Why does work cause fatigue? A real-time investigation of fatigue, and determinants of fatigue in nurses working 12-hour shifts.' *Annals of Behavioural Medicine,* 53(1); 1–12.
64. Jones, G. R., Deakin, A. G. and Spencer, J. W. (2005) 'Multi-stimulus Chromatic Processing of Complex Signals', *Proceedings of the complex systems monitoring session of the international complexity, Science and society conference*. Liverpool, UK, 11-14 September.
65. Jones, G. R., Deakin, A. G. and Spencer, J. W. (2008), *Chromatic monitoring of complex conditions,* Taylor & Francis Group: Boca Raton.
66. Jones, G. R., Russell, P. C., Vourdas, A., Cosgrave, J. Stergioulas, L. and Haber, R. (2000) 'The Gabor Transform Basis of Chromatic Monitoring.' *Measurement Science and Technology*, 11, pp. 489-498.
67. Kaida, K., Takahashi, M., Akerstedt, T., Nakata, A., Otsuka, Y., Haratani, Tn and Fukasawa, K. (2006) 'Validation of the Karolinska Sleepiness scale against performance and EED variables.' *Clinical Neurophysiology*, 117, pp. 1574-1581.
68. Kaliyaperumal, D., Elango, Y., Alagesan, M. and Santhanakrishanan, I. (2017) 'Effects of sleep deprivation on the cognitive performance of nurses working in shift.' *Journal of Clinical and Diagnostic Research,* 11, pp. 1-3.

69. Khanade, K. and Sasangohar, F. (2017) ‘Stress, Fatigue, and Workload in Intensive Care Nursing: A Scoping Literature Review’, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61, pp. 686–690.
70. Knipling, R. and Wierwille, W. (1994) ‘vehicle-based drowsy driver detection: current status and future prospects.’ DOI:10.1037/e526702009-001
71. Knutsson, A. (2003) ‘Health disorders of shift workers’ *Occupational medicine,* 53, pp. 103-108.
72. Koh, A., Dean, E., Zhang, I., Jones, G. R. and Spencer, J. W. (2005) ‘Effect of chromatic filter characteristics in quantifying complex data.’ *A chromatic approach to complexity, proceedings of the complex systems, monitoring session of the international complexity, science and society conference, Liverpool. CIMS centre ofr intelligent monitoring systems.*
73. Koh, A., Jones, G. R., Spencer, J. W. and Thomas, I. (2007) ‘Chromatic analysis of signals from a driver fatigue monitoring unit.’ *Measurement science and technology,* 18, pp.747-754.
74. Lerman, S., Mollicone, D. and Coats, S. (2017) ‘Use of the psychomotor vigilance test in fitness for work assessments.’ *journal of occupational and environmental medicine,* 59(8), pp. 716-720.
75. Lu, P. W., Columbus, A. B., Fields, A. M., Melnitchouk, N. and Cho, N. L. (2020) ‘Gender Differences in Surgeon Burnout and Barriers to Career Satisfaction: A Qualitative Exploration.’ *Journal of Surgical Research,* 247, pp. 28-33.
76. Ma , J., Lowndes, B., Chrouser, K., Hallbeck, S. M. and McCrory, B. (2021) ‘Developing a subjective instrument for laparoscopic surgical workload in a high fidelity simulator using the NASA_TYLX and SURG-TLX.’ *IISE transactions on healthcare systems engineering,* 11(2) pp. 161-169.
77. Mahdiani, S., Jeyhani, V., Peltokangas, M. and Vehkaoja, A. (2015) ‘Is 50hz high enough ECG sampling frequency for accurate HRV analysis?’ Conference: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society At: Milano, Italy
78. Mansukhani, M. P., Kolla, B. P., Surani, S., Varon, J. and Ramar, K. (2012) ‘Sleep deprivation in resident physicians, work hour limitations, and related outcomes: a systematic review of the literature.’ *Postgraduate medicine,* 124(4) pp. 241-249.

79. MATLAB. (2020). *version 9.8.0.1323503 (R2020a)*. Natick, Massachusetts: The MathWorks Inc.
80. Mayo Clinic (2023) *Heat Exhaustion*. Available at: https://www.mayoclinic.org/diseases-conditions/heat-exhaustion/symptoms-causes/syc-20373250 (Accessed: 5 January 2023).
81. McCormick, F., Kadzielski, J., Landrigan, C. P., Evans, B., Herndon, J. H. and Rubash, H. E. (2012) 'A prospective analysis of the Incidence, Risk, and Intervals of Predicted Fatigue-Related Impairment in Residents.' *The archives of surgery*, 147(5), pp. 430-435.
82. McKnight, P. E. and Najab, J. (2010) 'Mann-Whitney U Test' *The Corsini encyclopedia of psychology*, pp. 1-1.
83. Microsoft Corporation. (2016). *Microsoft Excel*. Retrieved from https://office.microsoft.com/excel
84. Miro, E., Cano-Lozano, M. C. and Buela-Casal, G. (2002) 'Electrodermal activity during sleep deprivation and its relationship with other activation and performance measures.' *Journal of Sleep Research*, 11(2) pp. 105-112.
85. Moore, D. S. and McCabe, G. P. 2003 *Introduction to the practice of statistics* (4th ed.). W. H. Freeman: New York.
86. Nejati, A., Shepley, M. and Rodiek, S. (2016) 'A review of design and policy interventions to promote murses' restorative breaks in health care workplces.' *Workplace health and safety*, 64(2), pp. 70-77.
87. Niu, S-F., Chung, M-H., Chen, C-H., Hegney, D., O'Brien, A. and Chou, K-R. (2011) 'The effect of shift rotation on employee cortisol profile, sleep quality, fatigue and attention level.' *Journal of nursing research*, 19(1) pp. 68-81.
88. Nurit, W. and Michal, A. (2003) 'Rest: A qualitative exploration of the phenomenon.' *Occupationalk therapy international*, 10, pp. 227-238.
89. Nussinovitch, U., Elishkevitz, K. P., Nussinovitch, M., Segev, S., Volovitz, B. and Nussinovitch, N. (2011) 'Reliability of ultra-short ECG indices for heart rate variability.' *Annals of Noninvasive Electrocardiology*, 16, pp. 17–22.
90. O'Callaghan, F., Muurlink, O. and Reid, N. (2018) 'Effects of caffeine on sleep quality and daytime functioning.' *Risk management and healthcare policy*. 11 pp. 263-271.

91. Or C. K. and Duffy, V.G. (2007) ‘Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement’, *Occupational Ergonomics*, 7(2), pp. 83-94.
92. Oriyama, S., Miyakoshi, Y. and Kobayashi, T. (2014) ‘Effects of two 15-min naps on the subjective sleepiness, fatigue and heart rate variability of night shift nurses.’ *Industrial Health.* 52(1) pp. 25-35.
93. Park, A., Lee, G., Seagull, F. J., Meenaghan, N. and Dexter, D. (2010) ‘Patients benefit while surgeons suffer: an impending epidemic.’ *Journal of American college of surgeons*, 210(3), pp. 306-313.
94. Park, A., Lee, G., Seagull, F. J., Meenaghan, N. and Dexter, D. (2010) ‘Patients benefit while surgeons suffer: an impending epidemic.’ *Journal of American college of surgeons*, 210(3), pp. 306-313.
95. Pichot, V., Bourin, E., Roche, F., Garet, M., Gaspoz, J-M., Duverney, D., Antoniadis, A., Lacour, J-R. and Barthelemy, J-C. (2002) ‘Quantification of cumulated physical fatigue at the workplace.’ *Pflugers Archiv: European journal of physiology*, 445, pp. 267-272.
96. Raudenbus, S. W. and Bryk, A. S. 2002. *Hierarchical linear models: Applications and data analysis methods* (vol. 1). Sage.
97. Reyner, L. A. and Horne, J. A. (1997) ‘Suppression of sleepiness in drivers: combination of caffeine with a short nap.’ *Psychophysiology*, 34(6), pp. 721-725.
98. Reyner, L. A. and Horne, J. A. (1998) ‘Falling asleep whilst driving: are drivers aware of prior sleepiness? *International journal of legal medicine*, 111, pp.120-123.
99. Roelands, B., Cutsem, Van Cutsem, J., Marcora, S. and Meeusen, R. (2016) ‘Does mental fatigue alter core and skin temperature in the heat?’ *Medicine and science in sport and exercise*, 48(5S), pp. 123.
100. Salahuddin, L., Cho, J., Jeong, M. G. and Kim, D. (2007) ‘Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings.’ *Conf Proc IEEE Engineering of Medicine and Biology Society*, pp. 4656–4659.
101. Santini, F., Onoratti, F., Faggian, G. and Mazzucco, A. (2011) ‘Surgeon sleep deprivation and outcomes in cardiac surgery: common sense, machismo, and statistics.’ *Archive of surgery*, 146(12) pp. 1453.

102. Shaffer, F. and Ginsberg, J. P. (2017) ‘An overview of heart rate variability metrics and norms.’ *Frontiers in public health*, 5(258), pp. 1-17.

103. Shahid, A., Wilkinson, K., Marcu, S. and Shapiro, C. M. 2011, *Karolinska Sleepiness Scale. STOP*, *THAT* and one hundred other sleep scales. Springer, New York.

104. Shanafelt, T. D., Balch, C. M., Bechamps, G. J., Russell, T., Dyrbye, L., Satele, D., Collicott, P., Novotny, P. J. Sloan, J. and Freischlag, J. A. (2009) ‘Burnout and career satisfaction among American surgeons.’ *Annals of surgery*, 250(3) pp. 463-471.

105. Simmonet, M. and Gourvennec, B. (2016) ‘Heart rate sensors acceptability: data reliability vs ease of use.’ *IEEE 13th international conference on wearable and implantable body sensor networks (BSN)*, pp. 94-98.

106. Stucky, C-C. H., Cromwell, K. D., Voss, R. K., Chiang, Y-J., Woodman, K., Lee, J. E. and Cormier, J. N. (2018) ‘Surgeon symptoms, strain and selection: Systematic review and meta-analysis of surgical ergonomics.’ *Annals of medicine and surgery*, 27, pp. 1-8.

107. Sturm, L., Dawson, D., Vaughan, R., Hewett, P,. Hill, A. G., Graham, J. C. and Maddern, G. J. (2011) ‘Effects of fatigue on surgeon performance and surgical outcomes: a systematic review.’ *ANZ journal of surgery*. 81(7-8), pp. 502-509.

108. Sufian, A. T., Jones, G. R., Shabeer, H. M., Elzagzoug, E. Y. and Spencer, J. W. (2018) ‘Chromatic techniques for in-vivo monitoring jaundice in neonate tissues.’ *Physiological measurement*, 39(9), 095004

109. Thomas, J. Q. (2016) One way analysis of variance (ANOVA) pp. 165-182.

110. Turhan, N. S. (2020) ‘Karl Pearson’s chi-square test.’ *Educational research and reviews*, 15(9) pp. 575-580.

111. Ulises, T., Hallowell, M., Stambaugh, N. and Littlejohn, R. (2016). ‘Causes and Consequences of Occupational Fatigue.’ *Journal of Occupational and Environmental Medicine*, 58(10), pp. 961-973.

112. Umair, M., Chalabianloo, N., Sas, C. and Ersoy, C. (2021) ‘HRV and stress: a mixed-method approach for comparison of wearable heart rate sensors for biofeedback.’ *IEEE Access*, 9, pp. 14005-14024.

113. Van Dongen, H. P. A., Maislin, G., Mullington, J. M. and Dinges, D. F. (2003) 'the cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation.' *SLEEP, 2, pp. 117-126.*

114. Van Watendonk, I., Smits, M., Merten, H., Heetveld, M. J. and Wagner, C. (2010) 'Nature, causes and consequences of unintended events in surgical units.' *British journal of Surgery*, 97(11) pp. 1730-1740.

115. Voss, R. K., Chiang, Y. J., Cromwell, K. D., Urbauer, D. L., Lee, J. E., Cormier, J. N. and Stucky, C. H. (2016) 'Do no harm, except to ourselves? A survey of symptoms and injuries in oncologic surgeons and pilot study of an intraoperative ergonomic intervention.' *Journal of the American College of Surgeons*, 244(1) pp. 16-24.

116. Voss, R. K., Chiang, Y. J., Cromwell, K. D., Urbauer, D. L., Lee, J. E., Cormier, J. N. and Stucky, C. H. (2016) 'Do no harm, except to ourselves? A survey of symptoms and injuries in oncologic surgeons and pilot study of an intraoperative ergonomic intervention.' *Journal of the American College of Surgeons*, 244(1) pp. 16-24.

117. Wang, X., Li, X., Rong, M., Xie, D., Ding, D. and Wang, Z. (2017) 'UHF signal processing and pattern recognition of partial discharge in gas-insulated switchgear using chromatic methodology.' *Sensors*, 17(177), pp. 1-15.

118. Wang, Z., Wang, X. H., Rong, M., Jones, G. R., Spencer, J. W., Humphries, J. E. and Yan, J. D. (2016) 'An optical study of contact erosion in a high voltage circuit breaker using a chromatic method.' *21st international conference on gas discharges and their applications*,

119. Wetter, D. W. and Young, T. B. (1994) 'The relation between cigarette smoking and sleep disturbance.' *Preventative medicine.* 23(3) pp. 328-334.

120. Wilson, M. R., Poolton, J. M., Malhotra, N., Ngo, K., Bright, E. and Masters, R. S. W. (2011) 'Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX).' *World journal of surgery*, 35(9), pp. 1961-1969.

121. Wu, D., Gross, B., Rittenhouse, K., Harnish, C., Mooney, C., Rogers, F. B. (2017) 'A preliminary analysis of compassion fatigue in a surgeon

population: Are female surgeons at heightened risk?' *The American Surgeon*, 83, pp.1032-1037.

122. Zhang, J., Jones, G. A., Deakin, A. G. and Spencer, J. W. (2004) 'Chromatic processing of DGA data produced by partial discharges for the prognosis of HV transformer behaviour.' *Measurement Science and Technology*, 16(2), pp. 556

Appendix A: Statistical tests used for thesis

Chi-squared test (Turhan, 2020)

The Chi-squared test is applied to categorical data to evaluate the likelihood that any observed differences between data sets arose by chance (Turhan, 2020). A *P* value <.05 was used for statistical significance. The formula for χ^2 is as follows:

$$\chi^2 = [O - E]^2/E$$

Where *O* is the observed frequency and *E* is the expected frequency (Turhan, 2020). The format for reporting includes the degrees of freedom, test statistics and p value related to the test statistic. Degrees of freedom is calculated as (r-1)(c-1), where r = rows and c = columns of data (rows = male and female, columns = yes and no). The format is displayed below, where; *df* = degrees of freedom, *N* = number of participants, *V* = value of χ^2 and *p* = statistical significance:

$$\chi^2 = (df\ ,\ N) = V,\ p$$

The percentage of answers for females and males are also reported, respectively.

Non-parametric assumptions and tests (Moore and McCabe, 2003)

Where data breached parametric assumptions, non-parametric tests were conducted. Typical assumptions for parametric data include normality, homogeneity of variance and outliers. These assumptions are explained below:

- Normality: Normality refers to the normal distribution of the data, that when visualised shows a 'bell-shaped' curve, implying the data is symmetrical and evenly distributed around the mean. Normality plots can be output in SPSS. The Shapiro-Wilk test of normality is used for sample sizes <50. If the significance

of the Shapiro-Wilk test is greater than 0.05, the data is determined to be normal. However if it is below 0.05, the data significantly deviate from the normal distribution.

- Homogeneity of variance: Certain tests require that the variance of the populations are equal. The Levene's Test is used to determine whether two or more groups have equal variance. A significant result (less than 0.05) indicates the violation of assumed homogeneity of variance.

- Parametric assumptions assume there are no extreme outliers in any group that could adversely affect the test result. Boxplots can visually check for outliers to see whether there are many results larger than the rest of the observations in the group.

In instances where at least one of these assumptions are not met, non-parametric tests are conducted. These are used when the data is skewed. Non-parametric tests are applicable to all types of data and include short calculations. However, they are less efficient compared to parametric tests and the results should be analysed cautiously as they may not provide an accurate answer due to the lack of distribution among the data.

The two non-parametric tests used in this body of work are Mann-Whitney U and Kruskal Wallis tests. They are described below.

Mann-Whitney U (McKnight and Najab, 2010): This test compares the outcomes of two independent samples. It is the non-parametric equivalent of the t-test. It is used to test whether observations in one sample tend to be larger than observations in the other. The significance level $a = 0.05$ is chosen. Firstly, ranks are assigned from both groups in order from smallest to largest. The test statistic can then be calculated. The equation for Mann-Whitney U is below:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

Where and n_1 and n_2 are the number of participants and R_1 and R_2 are the sum of the ranks in the two groups, respectively. For this test, the null hypothesis predicts the sum of the rankings in the two group does not differ (U_1 and U_2). The alternative population shows the sum of the rankings differs between the two groups.

$$U = \min(U_1, U_2)$$

$$\text{Expected U} \rightarrow \mu U = \frac{n_1 n_2}{2}$$

$$\text{Standard error of U} \rightarrow \delta U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

$$\text{Z-value} \rightarrow z = \frac{U - \mu U}{\delta U}$$

A Mann-Whitney table then cross-references the Z-value to calculate the p-value. The table provides critical values of *U* for various alpha and sixes of two samples for the two-tailed test used in this research (Real-Statistics, 2023). If the corresponding number in the table is greater than the Z=value, the null hypothesis is rejected ($p \leq 0.05$).

Kruskal-Wallis test (Moore and McCabe, 2003): The Kruskal-Wallis rest is used when there are three or more independent samples. The parametric equivalent is the One-way ANOVA. The test determines whether the medians of three or more groups are different. The significance level is *a = 0.05*. Firstly, the data from the three groups is sorted into ascending order in one combined set and assigned ranks. The ranks for each sample are calculated by adding up the rank numbers. The H statistic is then calculated with the equation below:

$$H = \left[\frac{12}{n(n+1)} \sum_{j=1}^{c} \frac{T_j^2}{n_j} \right] - 3(n+1)$$

Where n = sum of sample sizes for all samples, c = number of samples, T_j = sum of ranks in the j^{th} sample and n_j = size of the j^{th} sample. A chi-square value is found using the chi-square table (Statistics How To, 2023), with c-1 degrees of freedom. The H value is compared to the critical chi-square value. If the critical value is less than H, the null hypothesis can be rejected. If the chi-square value is not less than H, there is insufficient evidence suggesting the medians are unequal.

Linear Mixed Model (Raudenbush and Bryk, 2002)

Linear Mixed Models (LMMs) are an extension to typical linear models, such as linear regression, and they are often used with more complicated sampling designs.
Simple linear regression uses the following equation:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Where variable Y can be modelled by a dependent variable X with coefficients indicating the slope of the line and intercept on the X axis.

Multiple linear regression can be defined with the following equation, where multiple terms are integrated into the model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Regression models are suitable in data where the sample group is homogenous. The LMM however, considers experimental designs with multiple subgroups within a larger population. includes both random effects and fixed effects. The mixed effects model equation is defined below:

$$yij = (\beta 0 + b0i) + (\beta 1 + b1i)tij + \varepsilon ij$$

Where $b0i$ and $b1i$ are the random effects (one pair per subject), $b0i$is the subject-specific deviation from the average intercept, $b1i$ is the subject-specific deviation

from the average slope, $\beta 0$ and $\beta 1$ are the fixed effects, tij is the vector of time and εij is the vector of errors for observations in groups (Andrinopoulou et al., 2012).

Pearson Correlation (Freedman et al., 2007)

Pearson correlation coefficient analysis identifies the linear relationship between two variables. The output varies from -1 to 1, where -1 is a strong negative correlation and 1 is a strong positive correlation. A negative correlation refers to the change in one variable causing the other variable to change in the opposite direction. A correlation of 0 shows no relationship between the variables. A positive correlation shows both variables change in the same direction. The Pearson correlation coefficient describes the strength and direction of the linear relationship between two quantitative variables. The coefficient is given the value *r*. Generally, $r > .5$ is a strong correlation and $r > .7$ is a very strong correlation. Scatter plots can depict the visual relationship between two variables.

The relationship between two variables measures the strength in relationship of the samples only. It is important to draw conclusions about the populations, not just the samples. Therefore, it is required to do statistical significance tests. A hypothesis test is required, starting with the selection of null and alternative hypothesis. The alternative hypothesis is what is aimed to be achieved, whereas the null hypothesis is what is trying to be disproved. The hypothesis test aims to provide evidence that there is a significant correlation between the two variables. The standard measure used to measure significance is the *p*-value. The *p*-value is between 0 and 1, the probability that the null hypothesis is true. The value represents the probability that the correlation between two variables occurred by chance. The accepted *p*-value is ≤ 0.05, which indicates that there is only a $\leq 5\%$ chance that the results from the sample occurred due to chance. In this instance, the null hypothesis can be rejected in favour for the alternative which concludes the correlation is statistically significant. If the *p*-value is bigger than the significance level, there is a failure to reject the null hypothesis and the correlation is not statistically significant.

Correlation measures the direction and strength of the relationship between two variables but regression predicts the value of variable *Y* based on the known value of the variable *X* through the following equation:

$$Y = a + bX$$

Where b indicates the slope of the line and a is the intercept (the value of *Y* when *X* = 0).

Pearsons correlation is reported as:

r(degress of freedom) = the *r* statistic, *p* = *p* value.

Degrees of freedom (df) represent the number of values in the final calculation of a statistic that are free to vary. In other words, it's the number of values in the final calculation that can be chosen independently.

The correlation coefficient (often referred to as "r") quantifies the strength and direction of a linear relationship between two continuous variables. It ranges from -1 to +1.

Regression Correlation (Freedman et al., 2009)

Regression analysis is analysed using the R^2 statistic, which represents the proportion of variance for a dependant variable which is explained by an independent variable. The dependant variable is the factor trying to be understood or predicted, in this instance, KSS. The independent variable is a factor hypothesised to have an impact on the dependent variable. Whilst correlation explains the strength of the relationship between the independent and dependant variables, R^2 explains the extent the variance of one variable explains the variance of a second variable. This can be written by the following equation:

$$R^2 = 1 - \frac{\text{Unexplained variation}}{\text{Total variation}} \qquad \text{---------- Eqn 8}$$

If the R^2 value of a model is 0.60, then approximately 60% of the observed variation can be explained by the model's inputs. Calculating the R^2 requires data points of dependant and independent variables and finding the line of best fit. A 'good' R^2 value

is flexible depending on its application. Social research fields generally accept 0.5, however a finance field may accept 0.7 as a strong value.

T-test (Freedman et al., 2009)

A T-test is used to compare two groups which meet the parametric assumptions. It is often used to see whether two groups are different from one another. An independent-samples T-test investigates whether there is a difference between two groups (between-subjects). In this work, a two-tailed t-test is carried out as the difference on populations is calculated. The formula for the T-test is below:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s^2(\frac{1}{n_1} + \frac{1}{n_2}))}}$$

In this formula, t is the t value, x_1 and x_2 are the means of the two groups being compared, s_2 is the pooled standard error of the two groups, and n_1 and n_2 are the number of observations in each of the groups. A larger t value shows that the difference between group means is greater than the pooled standard error, indicating a more significant difference between the groups. The calculated t value is compared against the values in a critical value chart to determine whether the t value is greater than what would be expected by chance. If so, the null hypothesis is rejected, and concludes that the two groups are in fact different. The formula for displaying the results is as follows:

t(degrees of freedom) = t value, p = p value

Analysis of Variance (ANOVA) (Moore and McCabe, 2003)

The analysis of variance is used to compare the means of three or more independent samples, which meet parametric assumptions. An ANOVA is used instead of multiple

T-tests as multiple T-tests increases the chances of a type I error (false positive) as the hypothesis cannot be tested simultaneously, thus increasing the p value. A One-Way ANOVA refers to the number of independent variables, in this case there is one. The means of the independent groups are compared using the F-distribution. The F value can report significant differences between the levels of the independent variable, when $p < 0.05$. A higher F value indicates the means are more significantly different. The F value equation is below:

$$F = \frac{MST}{MSE}$$

$$MST = \frac{\sum_{i=1}^{k}(T_i^2/n_i) - G^2/n}{k-1}$$

$$MSE = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^{k}(T_i^2/n_i)}{n-k}$$

Where F is the variance ratio for the overall test, MST is the mean square due to treatments/groups (between groups), MSE is the mean square due to error (within groups, residual mean square), Y_{ij} is an observation, T_i is a group total, G is the grand total of all observations, n_i is the number in group i and n is the total number of observations. The formulae for writing ANOVA is:

F(between groups df, within groups df) = [F-value], p = p-value.

When the initial F test indicates that significant differences exist between group means, post hoc tests are useful for determining which specific means are significantly different when specific hypotheses are not tested. Post hoc tests compare each pair of means but unlike t-tests, they correct the significance estimate to account for the multiple comparisons.

Appendix B – surgeon questionnaire

Fatigue in the Workplace

1. Surgeon Questionnaire

The idea of this anonymous questionnaire is to try and identify the prevalence of workplace fatigue, understand the onset and any interventions you may use to counteract fatigue. Your participation is greatly appreciated and should take no more than 10 minutes to answer 15 questions.
Please note that your data will be restricted to the research team and aims to be untraceable. The data will be handled with care and aims to be reported in a journal identifying aspects of fatigue in the workplace. The University processes personal data as part of its research and teaching activities in accordance with the lawful basis of 'public task', and in accordance with the University's purpose of "advancing education, learning and research for the public benefit." Under UK data protection legislation, the University acts as the Data Controller for personal data collected as part of the University's research.
The Principal Investigator acts as the Data Processor for this study, and any queries relating to the handling of your personal data can be sent to [Jacob Merriman (j.merriman@liverpool.ac.uk) / Supervisor: Jim Humphries (humph9ke@liverpool.ac.uk)]. Please proceed with the questionnaire if you agree with the following statements:

I confirm that I have read and have understood the information sheet for the study, or it has been read to me.

I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

I understand that my participation is voluntary and that I am free to stop taking part and can withdraw from the study at any time without giving any reason and without my rights being affected.

In addition, I understand that I am free to decline to answer any particular question or questions. I understand that the information I provide will be held securely and in line with data protection requirements at the University of Liverpool which is fully anonymous and then deposited in the archive for sharing and use by other authorised researchers to support other research in the future.

I agree to take part in the above study.

1. What is your gender?

☐ Male

- [ ] Female
- [ ] Prefer not to say
- [ ] Other (please specify):

2. What is your age range?

- [ ] 25-35
- [ ] 36-45
- [ ] 46-55
- [ ] 56-65
- [ ] 66-75
- [ ] 76+
- [ ] Prefer not to say

3. What is your duration of surgical experience?

- [ ] <5 years
- [ ] 5-14 years
- [ ] 14-25 years
- [ ] 25-34 years
- [ ] 35-44 years
- [ ] 45+ years
- [ ] Prefer not to say

4. Please state your department of work and Surgical Grade?

☐ Prefer not to say

My surgical department is:

5. Would you say you have been subject to any of the following sources of fatigue in the last month? (Choose all that apply)

☐ Musculoskeletal fatigue (aches/pains in the muscles/joints as a result of surgery)

☐ Sleep deprivation (the desire to sleep/nap whilst at work due to lack of sleep)

☐ Physical exhaustion (lack of energy to complete day-to-day tasks)

☐ Mental exhaustion (lack of desire/willingness to execute a task)

☐ Compassion fatigue (emotionally drained as a result of frequency of suffering patients and/or friends/relatives of patients)

☐ Drowsiness whilst driving (fighting the urge to sleep whilst driving to/from work)

☐ I have not been subject to fatigue at work

☐ Prefer not to say

☐ Other (please specify):

6. What contributes to your fatigue at work/ whilst performing surgery? (Choose all that apply)

☐ I have not experienced fatigue at work

- Persistent mental activity
- Prolonged static posture
- Continual tension using hands
- Monitor height
- Bright lights
- Lack of control at work
- Patient outcomes
- Lack of motivation
- High levels of effort
- Lack of rest periods during shift
- Lack of sleep
- Physical exertion
- Table height
- Positioning of apparatus
- Uncomfortable posture
- Length of operation
- Discussion with friends/families of patients

- [ ] High levels of demand
- [ ] Disputes with colleagues
- [ ] Lack of reward/recognition
- [ ] Lack of food/water during shift
- [ ] Prefer not to say
- [ ] Other (please specify):

7. How have you attempted to minimise these problems/conditions at work/ during surgery? (Choose all that apply)

- [ ] Adjust the surgical field
- [ ] Change instruments
- [ ] Take a break whenever conveniently possible
- [ ] Change position
- [ ] Discussion with supervisors
- [ ] Relaxation/exercise techniques (stretches, deep breaths, etc.)
- [ ] Change height of surgical field
- [ ] Treatment/medication
- [ ] Ignore it
- [ ] Time away from operating room
- [ ] 'It's impossible', 'it's part of the job'
- [ ] I have not needed to minimise fatigue at work

- [ ] Prefer not to say
- [ ] Other (please specify):

8. Thinking about your average work week, how many hours are you at work?

- [ ] <20
- [ ] 20-30
- [ ] 30-40
- [ ] 40-50
- [ ] 50-60
- [ ] 60-70
- [ ] 70+
- [ ] Other (please specify):

9. Thinking about your average work week, how many hours do you sleep per night?

- [ ] <2
- [ ] 2-4
- [ ] 4-6
- [ ] 6-8
- [ ] 8-10
- [ ] 10+

10. Have you ever had to miss work/ leave work early as a result of fatigue?

- [ ] Yes
- [ ] No
- [ ] Unsure
- [ ] Prefer not to say
- [ ] Other (please specify):

11. Do you use or are you aware of any intervention strategies to mitigate workplace fatigue?

- [ ] I am not aware of any interventions to reduce fatigue but I am subject to fatigue
- [ ] I am aware of interventions to reduce fatigue but do not use any myself
- [ ] I am not aware of any interventions to reduce fatigue as I do not need to know any
- [ ] Prefer not to say

I am aware of interventions to reduce fatigue and my intervention(s) is:

3. Microbreaks

Current research is formulating around the idea of taking short pauses during operations to help reduce the effects of musculoskeletal fatigue in surgeons whilst performing long operations. 'Microbreaks' (~1 minute pauses every 20-40 minutes with opportunity to stretch, relax and communicate) have been shown to improve physical and mental well-being of surgeons without any prolonged operation duration or adverse patient outcomes.

12. What are your thoughts on 'Microbreaks' as a possible intervention?

- [ ] I would be willing to see whether they reduced my fatigue symptoms
- [ ] I am hesitant about the idea of taking structured breaks during surgery

- [ ] They seem like a viable option
- [ ] I would need more information on the 'Microbreaks' before I can make a decision
- [ ] They do not seem advantageous to me
- [ ] Taking breaks during surgery would disrupt my flow
- [ ] I would opt for an alternative solution
- [ ] Prefer not to say
- [ ] Other (please specify):

Comments:

4. Wearables

Wearable technology is on the rise (such as watches which can monitor your heart rate), with new electronics being able to monitor vitals throughout the working day. As of yet, there is no wearable suitable for the surgical environment to predict the onset of fatigue at the individual level. Ideas are postulating around a possible wearable vest to be worn during the shift, that when an individual is to become fatigued it can suggest that a break is required.

13. What are your thoughts on the idea of a base-layer wearable vest monitoring your fatigue levels during your shift?

- [ ] I would wear the vest to monitor my fatigue levels
- [ ] I do not need to wear the vest
- [ ] I would need more information on the vest before I can make a decision
- [ ] I would not wear the vest to monitor my fatigue levels
- [ ] I am hesitant about its benefits, however I would give it a go
- [ ] Prefer not to say

- [ ] Other (please specify):

Comments:

14. If you find yourself fatigued after a shift, does this influence your decision on whether you should drive home?

- [ ] I do not drive to/from work
- [ ] Yes - I usually take a break/nap before driving home
- [ ] No - I usually drive straight home
- [ ] I have not been fatigued at the end of my shift before getting into my car
- [ ] Yes - I am aware that I should not be driving however I have errands to run
- [ ] Unsure
- [ ] Prefer not to say
- [ ] Other (please specify):

Comments:

15. Would you be willing to use a self-management application on your mobile phone to monitor your fatigue levels?

- [ ] Yes
- [ ] No

- [ ] Unsure
- [ ] Prefer not to say

Appendix C

Example of MATLAB code

```
%Load the file. Format is P(number of participant_hour_minute_second)

peaks = load('PX_XX_XX_X.txt');
max = 320;
min = 0;
dif = diff(peaks);
N = (max-min)./(dif-min);
BPM = N*60;
L = length(BPM);
U = round(L/15);



%works out average BPM every minute (split into 5 equal sections)

%if we have 450 samples, this breaks it up into 5 equal sections (450/90=5)

blockSize = [U, 1];

meanFilterFunction = @(theBlockStructure) mean2(theBlockStructure.data(:));

Block = blockproc(BPM, blockSize, meanFilterFunction);

% Let's check the output size.

%Now we can calculate HRV measures.
```

```
[rows, columns] = size(Block);


B=60000;
A=BPM;
%convert each HR sample to ms
X = B./A;


RMSSD = sqrt(X);
%works out SDNN every minute
X(isinf(X)) = 0;
X(X==0) = [];


blockSize1 = [U, 1];
meanFilterFunction = @(theBlockStructure) mean2(theBlockStructure.data(:));
% % % % Now do the actual averaging (block average down to smaller size array).


Block1 = blockproc(RMSSD, blockSize1, meanFilterFunction);
% % % % Let's check the output size.
[rows, columns] = size(Block1);
```

Example of Microsoft Excel for the calculation of the Predicted KSS variable

	maximum parameter	7								
	mimimum parameter	1			Circadian Rhythm					
Time	Normalised	Raw value	R	G	B	Max	Min	Saturation	Lightness	Hue
8:00:00	0	1								
8:00:01	0.0004167	1.0025								
8:00:02	0.0008333	1.005								
8:00:03	0.00125	1.0075								
8:00:04	0.0016667	1.01	0.0075	0.015	0.0225	0.0225	0.0075	0.5	0.015	270
8:00:05	0.0020833	1.0125	0.01125	0.01875	0.02625	0.02625	0.01125	0.4	0.01875	270
8:00:06	0.0025	1.015	0.015	0.0225	0.03	0.03	0.015	0.3333333	0.0225	270
8:00:07	0.0029167	1.0175	0.01875	0.02625	0.03375	0.03375	0.01875	0.2857143	0.02625	270
8:00:08	0.0033333	1.02	0.0225	0.03	0.0375	0.0375	0.0225	0.25	0.03	270
8:00:09	0.00375	1.0225	0.02625	0.03375	0.04125	0.04125	0.02625	0.2222222	0.03375	270
8:00:10	0.0041667	1.025	0.03	0.0375	0.045	0.045	0.03	0.2	0.0375	270
...	...	...	...	...	...	...	...	...	...	...

This screen shot of Microsoft Excel uses the chromatic formulas in equations 3,4 and 5 in Chapter 2.4.3, for each row in the saturation, lightness and hue columns for the data depicted in the Raw value column.

Example of Three-channel ECG data

Timestamp	Ch1	Ch2	Ch3
0	4.936071	-1.61979	-3.11486
0.003125	4.935107	-1.62611	-3.11636
0.00625	4.936929	-1.63168	-3.11904
0.009375	4.933607	-1.64604	-3.12825
0.0125	4.92825	-1.67132	-3.144
0.015625	4.928893	-1.67111	-3.13618
0.01875	4.934893	-1.65921	-3.12525
0.021875	4.934357	-1.65129	-3.12225
0.025	4.931143	-1.65236	-3.12493
0.028125	4.932429	-1.66189	-3.12771
0.03125	4.929857	-1.67325	-3.13114
0.034375	4.927929	-1.67411	-3.13564
0.0375	4.923321	-1.67057	-3.14057
0.040625	4.921714	-1.66639	-3.14014
0.04375	4.930929	-1.66029	-3.13125
0.046875	4.936393	-1.66243	-3.12793
0.05	4.931679	-1.67164	-3.12879

Example of RRI peaks with ratio column for one channel of ECG

Rpeak_Ch1 (sample pts)	R_amp_Ch1 (mV)	R1/BL_Ch1	R2/BL_Ch1
88	1.046480046	11.25316456	11.26582278
254	1.055028487	10.68	10.156
425	1.000734188	14.10555556	14.40555556
595	1.015211912	13.16243655	13.69543147
767	1.048573293	12.78672986	13.01895735
935	1.061451967	11.30452675	11.19341564
1108	1.02703946	13.8071066	14.08121827
1278	1.093223076	6.391705069	6.223502304
1448	1.035576922	18.37414966	18.03401361
1616	1.017406187	5.107899807	5.95761079
1783	1.196626306	15.7755102	12.99489796
1950	1.018266043	12.36407767	12.50970874
2116	1.018898598	6.394540943	7.079404467
2282	1.209452455	12.79372197	11.17488789
2450	1.040434984	0.741005055	0.949152542
2617	1.042849955	1.995	1.72

Device specifications for chest strap

13. Device Specifications

Performance Characteristics

- ECG Channel ..... 3 Channels
- ECG Dynamic Range.....+/- 400mV
- Memory Storage Format..... 64Mb Flash Memory
- Pod Operational Time ..... 1 year warranty
- Garment Operational Time ..... 30 day warranty
- Frequency Range .....DC to 65 Hz
- CMRR.....80 - 100 dB
- Input Impedance ..... .> 1 GΩ
- Sampling Rate.....320 samples/second
- ADC Resolution ..... 24 bits DC
- Communication Interface Protocol .... Bluetooth Low Energy 4.2
- Operating Frequencies.....2.4-2.48GHz
- Frequency Deviation ..... 2 MHz/channel

Physical Characteristics

- Skiin Pod.....12.0 grams +/- 0.5g 27.9x46.5x10.6mm
- Skiin Underwear .....Men's and Women's sizes range from 24" to 47"

Power Requirements

- Battery Type .....Rechargeable Lithium Polymer
- Power Supply Model.....KA1803A-US (Ke Yu Dian Yuan (HongKong) International Ltd.)
- Battery Capacity .....200mAh, 2-3 days of use
- Battery Life ..... minimum 24 Hours
- Operational Time..... 1 year warranty

Environmental Specifications

- Operational Temperature .....5 to 40°C
- Operational Humidity..... 15% to 90%, non-condensing, water vapour partial pressure <50hPa
- Operational Altitude..... 10kPa to 106kPa
- Storage Humidity..... 15% to 90%, non-condensing, water vapour partial pressure <50hPa
- Storage Temperature ..... -10°C to 45°C (for 1 month), -10°C to 35°C (for 6 months), or -10°C to 28°C (over 6 months)

Appendix D

MATLAB code for frequency domain analysis

```
clear
clc

% Load raw ECG data
data = load('p1_freq_final_02_14.txt');

% Define sampling frequency (in Hz)
fs = 320;

% Create time axis based on the number of samples and the sampling
frequency
t = (0:length(data)-1)/fs;

% Plot raw ECG data with linear scaling on Y-axis
figure;
plot(t, data);
xlabel('Time (s)');
ylabel('Voltage (mV)');
title('Raw ECG Trace');

% Set the desired cutoff frequency for the high-pass filter
fc = 0.04;

% Set the filter order (a second-order Butterworth filter is a common
choice)
order = 2;

% Use the Butterworth filter design function to generate the filter
coefficients
% for a high-pass filter with the desired cutoff frequency and filter
order
[b, a] = butter(order, fc / (fs/2), 'high');

% Apply the filter to the ECG data using the filtfilt function, which
filters the
% signal forward and backward to eliminate phase distortion
ecg_f = filtfilt(b, a, data);

% Differentiate the ECG signal
ecg_diff = diff(ecg_f);

% Square the differentiated signal
ecg_diff_sq = ecg_diff.^2;

% Integrate the squared signal
N = round(0.15*fs);
b = ones(1,N)/N;
a = 1;
int_ecg_diff_sq = filter(b,a,ecg_diff_sq);

% Find R-peaks using the Hamilton-Tompkins algorithm
[~,r_i_raw] = findpeaks(int_ecg_diff_sq,'MinPeakHeight',0.0005,...
```

```
    'MinPeakDistance',0.35*fs,'MinPeakProminence',0.0005);

% Create a tachogram by calculating the time differences between
consecutive R-peaks
IBI = diff(r_i_raw)/fs*1000; % convert to milliseconds;
IBI_transposed = IBI';

% Plot results
figure;
plot(t,ecg_f);
hold on;
plot(t(r_i_raw),int_ecg_diff_sq(r_i_raw),'ro','MarkerSize',20);
xlabel('Time (s)');
ylabel('ECG (mV)');
legend('ECG signal','R-peaks');
 %}
    % Compute the Lomb-Scargle periodogram
    [p,f] = plomb(IBI_transposed,[],[],fs);

    % Compute the power spectral density
    psd = p/(length(IBI_transposed)/fs);

    % Define the frequency bands of interest
 %   ULF_band = [0 0.04]; % Hz
    LF_band = [0.04 0.15]; % Hz
    HF_band = [0.15 0.4]; % Hz

        % Find the indices of the frequency bands
    LF_indices = find(f >= LF_band(1) & f <= LF_band(2));
    HF_indices = find(f >= HF_band(1) & f <= HF_band(2));


% Divide the tachogram into 3 equal sections
num_sections = round((length(ecg_f))/(300*fs));
section_length = floor(length(IBI_transposed) / num_sections);

% Calculate the power spectral density for each section using a for
loop
psd_sections = zeros(num_sections, 1);
    lf_power1 = zeros(num_sections, 1);
    hf_power1 = zeros(num_sections, 1);
    a_lf_hf_ratios1 = zeros(num_sections, 1);

    %psdAreasCell = {};  % Initialize an empty cell array
    a_psdAreasMat = [];  % Initialize an empty matrix

for i = 1:num_sections
    section_start = (i-1)*section_length+1;
    section_end = i*section_length;
    if i == num_sections
        section_end = length(IBI_transposed);
    end
    section_IBI = IBI_transposed(section_start:section_end);
    [p,f] = plomb(section_IBI,[],[],fs);
    psd = p/(length(section_IBI)/fs);
    psd_sections(i) = mean(psd);

        % Compute the LF and HF power
    lf_power = sum(psd(LF_indices));
```

```
    hf_power = sum(psd(HF_indices));

    lf_power1(i) = sum(psd(LF_indices));
    hf_power1(i) = sum(psd(HF_indices));

    % Calculate the LF/HF ratio
    lf_hf_ratios = lf_power / hf_power;
    a_lf_hf_ratios1(i) = lf_power1(i)/hf_power1(i);

     figure();

    % Plot the power spectral density
    plot(f, psd, 'k'); % Black line
    hold on;

    % Highlight the LF and HF areas in different colors
    plot(f(LF_indices), psd(LF_indices), 'r', 'linewidth', 2); % Red
line
    plot(f(HF_indices), psd(HF_indices), 'b', 'linewidth', 2); % Blue
line

    % Add labels and legend
    xlabel('Frequency (Hz)');
    ylabel('Power spectral density');
    legend('PSD', 'LF', 'HF');
    xlim([0.04 0.4]);

    % Add shaded regions for each frequency band of interest
   % ulf_power = area(f, psd.*(f>=ULF_band(1)&f<=ULF_band(2)),
'FaceColor', [1 1 0.7], 'EdgeColor', 'none');
    lf_area = area(f, psd.*(f>=LF_band(1)&f<=LF_band(2)),
'FaceColor', [0.9 0.9 1], 'EdgeColor', 'none');
    hf_area = area(f, psd.*(f>=HF_band(1)&f<=HF_band(2)),
'FaceColor', [1 0.9 0.9], 'EdgeColor', 'none');
   %}
    a_psdAreas = [];
        fRange = 0.04:0.001:0.4;

    for j = 1:length(fRange)-1
        fIdx = find(f >= fRange(j) & f < fRange(j+1));
        a_psdAreas = [a_psdAreas, trapz(f(fIdx), psd(fIdx))];
    end

    % Concatenate the new row onto the existing matrix
    a_psdAreasMat = vertcat(a_psdAreasMat, a_psdAreas);
    a_PSD = a_psdAreasMat';
end
          %{
       for j = 1:length(fRange)-1
        fIdx = find(f >= fRange(j) & f < fRange(j+1));
        psdAreas = [psdAreas, trapz(f(fIdx), psd(fIdx))];
    end

    psdAreasCell{i} = psdAreas;  % Store the psdAreas values in the
cell array
end

% Convert the cell array to a matrix (if all arrays have the same
size)
```

```
psdAreasMat = cell2mat(psdAreasCell);

%}
```