Multi Day Fatigue Computation using Artificial Intelligence and a Single Sensor in an Uncontrolled Environment

Brian Russell

Bachelor of Engineering in Electrical and Electronics

A thesis submitted to the Auckland University of Technology
In fulfilment of the degree of Doctor of Philosophy

28 May 2020

Sport Performance Research Institute New Zealand
School of Sport and Recreation
Faculty of Health and Environmental Sciences
Auckland University of Technology

Primary Supervisor – Professor Patria Hume (AUT)
Secondary Supervisor – Associate Professor Andrew McDaid (UoA)
Tertiary Supervisor – Dr William Toscano (NASA)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ATTESTATION OF AUTHORSHIP

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institution of higher learning, except where due acknowledgement is made in the acknowledgements.

Chapters 3, 4 and 5 of this thesis represent separate papers that have either been published or have been submitted to peer-reviewed journals for consideration for publication. My contribution and the contributed by the various co-authors to each of these papers are outlined at the beginning of each chapter. All co- authors have approved the inclusion of the joint work in this doctoral thesis.

……………………..

Brian Russell

12th May 2020

CANDIDATE CONTRIBUTIONS TO CO-AUTHORED PAPERS

Chapter Publication reference	Author %
Chapter 3. Russell B.K., Hume P.A., McDaid A., Toscano, W., Calculating compliance for wearables and cognitive assessments in the field: A Case Study during a 4-person multi-day offshore sailing voyage, Journal of Sports Psychology in Action (submitted)	BR 87% PH 5% AM 5% WT 3%
Chapter 5. Russell B.K., Hume P.A., McDaid A., Toscano, W., **Moving the Laboratory into the Mountains: A Pilot Study of Human Activity Recognition In Unstructured Environments**, IEEE Transactions on Biomedical Engineering (submitted)	BR 85% PH 5% AM 5% WT 3%
Chapter 6. Russell B.K., Hume P.A., McDaid A., Toscano, W., **Predicting fatigue with a single sensor and machine learning in mountain terrain**, International Journal of Sports Physiology and Performance (submitted)	BR 87% PH 5% AM 5% WT 3%

Patria Hume, PhD	Andrew McDaid, PhD	Willian Toscano, PhD
Auckland University of Technology	University of Auckland	NASA Ames Research

Peer-reviewed conference poster and abstract

Brian Wang, Eleni Antoniadou, David Belo, Krittika D'Silva, Annie Martin, Brian Russell, Graham Mackintosh, Tianna Shaw and Frank Soboczenski, **Generative Models for Synthesizing Symptomatic ECG Astronaut Health Data for Future Deep Space Missions**, Second AI and Data Science Workshop for Earth and Space Sciences hosted at Jet Propulsion Laboratory, March 2020 (poster).

B Beard, B Russell, W Toscano, B Burian, M Krihak, S Shetye and T Shaw **Envisioning the Future Role of an Exploration Clinical Decision Support System**, Applied Human Factors and Ergonomics, conference track Neuroergonomic and Cognitive Engineering, San Diego, CA, July 2020 (abstract).

ACKNOWLEDGEMENTS

I would like to express my thanks to my primary supervisor, Prof. Patria Hume, who has continuously supported and encouraged my work through the course of my PhD, starting with the question of applying AI in the mountains to solve the question of fatigue and performance. She has challenged and guided my research directions making this process a significant learning experience. Also I would like to thank my co supervisors: Andrew McDaid for tirelessly supporting me on applying artificial intelligence to real world scenarios and keeping my work focused while allowing me to wander and learn; William Toscano for his wealth of knowledge in experimental human research in extreme environments; and Stacy Simms for her guidance in human systems modelling, extreme sport testing and experimental design.

I would also like to thank my colleagues at Auckland University of Technology, University of Auckland and NASA for their support, encouragement and discussions on various topics of discovery. In particular to Andrew Chen for guidance on Python programming for machine learning and Suzanne Barker-Collo for her help on understanding cognitive assessment differences between clinical and research contexts.

Lastly I would like to thank my wife, Nicole, and daughters Emily and Isabelle for supporting me through the process and listening to all the discoveries in field research and applying the amazing fields of physiology, biomechanics, neuropsychology, artificial intelligence in the mountains.

ETHICS

All experimental procedures in this study that involved human participants were subjected to ethical approval, which was endorsed by the Auckland University of Technology Ethics Committee. The thesis has two field research experiments. The first involved offshore sailing which was covered by AUTEC 17/353 and the second experiment, used for chapters 4 and 5, was covered by AUTEC 18/412.

AUTEC 17/353 Measurement of cognitive and physical performance during off-shore sailing.

AUTEC 18/412 Measurement of cognitive and physical performance during a 24-hour exercise test.

ABSTRACT

Multi-day field events such as adventure races in the mountains and military missions require executive decision making, resilience, endurance and physical performance. Cognitive fatigue and physical fatigue are significant factors in decision-making, physical performance and safety. Fatigue has been studied for over a century with work performed to understand mechanistic results in the laboratory. The literature published on fatigue has concluded that fatigue is a complex multi-variate problem with systems including physiology, neurology and psychology where the context of the field is required to be fully representative. There are challenges moving into the field including logistics, validation, battery power, equipment size, assessment distractions from the mission task, validation protocols. Machine learning incorporates a process of collecting data and training models to reach a desired prediction accuracy. The model performance is increased by either tuning the model or adding new data to cover new scenarios the model has not previously observed. Wearable technology is capable of gathering appropriate data with minimal distractions compared to traditional laboratory assessments and machine learning is capable of analysing vast amounts of data. The thesis investigated if an AI model could accurately predict cognitive and physical fatigue using a new dataset from a single sensor with field induced noise from an unstructured environment including obstacles and terrain variation.

In the narrative literature review, the questions included: What is the definition of fatigue?, What sensor data are available from various positions on the human body?, and what cognitive and physical assessments can be used to validate field work? The key findings were that fatigue is a result of complex overlapping systems and the definition of fatigue requires context. A subset of available sensors was possible when considered against the limitations imposed by remote field environments and exercising participants. There were various cognitive and physical assessments possible that could be applied to field environments which covered different types of cognitive performance deficits. Having identified the definition, sensing and validation mechanisms for cognitive and physical fatigue the next step was to perform a field study to assess the practicality of the protocol.

In chapter 2 the first of two field experiments is described for a 12 day offshore ocean sailing voyage. The questions asked included: what cognitive assessments are sensitive to field loads, what tools and protocol is practical in remote multi day environments. A formula was derived to determine the contributing factors to compliance of an assessment protocol. Findings from this field study included; an assessment should be a single software tool to aid the researcher, compliance to protocols is challenging in a field environment with fatigued participants, it is possible to formulate compliance and use this as an approach when designing protocols, field protocols take time in logistics, data collection and analysis such that pilot studies are imperative to prove validity and logistics before scaling the protocol to multiple participants.

A set of assessment and cognitive load software tools were developed based on the learnings from the offshore sailing study. These were to stream line the participants experience and also aid the researcher with logistical challenges such as tracking data during the protocol.

A new protocol was designed for the second field study with the new software tools. The protocol allowed an hourly cycle of field load and assessments to take place. The aim of the second study was to determine if classification of human activity recognition (HAR) was possible in the field with a single sensor. A trail run was used as the new study with previous reported lab results as the comparison study. To the authors knowledge, this is the first study in the field with a single sensor and AI model using noise sources such as terrain, fatigue and self-pacing. Findings included a trail calibration protocol and data pipeline to process the raw sensor data to enable an AI model to be trained. The AI model was optimised to trade-offs between repetitive activities such as running compared to one off actions such as climbing a fence. Results matched previously published results from controlled environments with no terrain or fatigue variation (accuracy 97.8% vs 97.7% for trail vs lab).

The last experiment was to determine if fatigue could be modelled in a field environment. A second AI model was developed for regression and trained against the cognitive and neuromuscular assessments that showed greatest sensitivity to the protocol, that being finger tap test (FTT) and vertical jump. It was confirmed that the

AI predictions using FTT had a mean absolute error (MAE) of 12.5% showing similar accuracy to laboratory based prior research.

In conclusion this thesis showed that fatigue is a complexity of redundant overlapping systems and a definition requires context from the field and mission goal to be applicable. Compliance is calculable and needs to be designed into the protocol to be successful in the field. A single sensor when coupled with a deep learning model can accurately classify human activity recognition. Fatigue can be predicted in the field using a single sensor and trained AI model. Further work is required to automate the data pipeline and protocol to enable multiple subject variation to be studied in the field.

1. INTRODUCTION AND RATIONALISATION

Research Motivation

For years I have used remote physiological and biomechanical systems to measure people performing tasks at the peak of their ability in challenging environments, from Fire Fighters in Puerta Rico with dehydration and heat stress to Special Forces operators being measured in shoot houses or planning tactical combat care. I have experienced some challenging situations with hypothermia mountain climbing, sleep deprivation sailing across a few oceans and dehydration and hypoxia flying a hang glider at altitudes above 13,000 feet. In all these situations people want assistance making decisions and performing to stay safe and successfully complete a mission. Injury from fatigue can occur due to a misjudged step or develop over time as the body compensates for wear and tear. However a recurring theme in remote physiological monitoring is that the operator only wants to carry equipment that helps with the mission and there is never enough time to look at a laptop screen to interpret engineering type plots. The medic or coach needs to be doing a job when there is action and long term trends are subtle and difficult to observe it the field. A metric often asked for when determining what is required to make all this technology and research valuable in the field is ‘give me a fuel tank indicator for the person and an ‘indicator of future performance’.

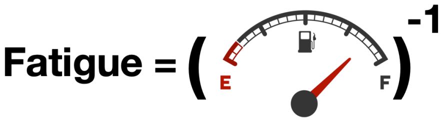


Figure 1-1 The relationship of fatigue and reserves

This translates into capacity to take on future work for a mission and the current level of performance at this point in time. The most influential factors on a person performing a task safely and competently is their physical ability and their mental ability to interpret a situation, plan and act. This has been explained by Boyd1 with the observe orient decide and act (OODA) loop.

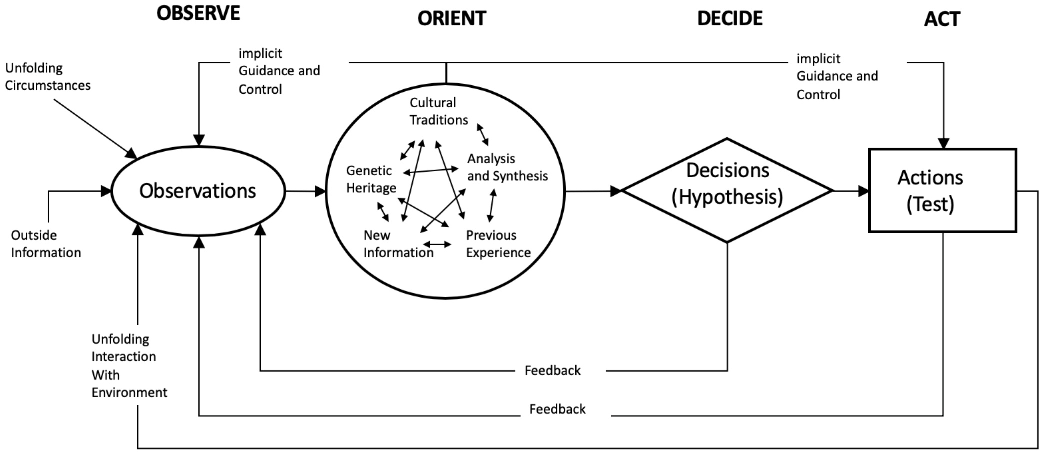


Figure 1-2 John Boyd's OODA Loop Diagram

Physical fatigue and cognitive fatigue limits a person’s ability to observe, orient, decide and act hence reduces their capacity to win.

Research Objectives

The research objective of the thesis was to determine if a fieldable solution of technology could predict fatigue with no burden of interpretation by implementing a sensor and deep learning model in the field.

This required the definition of fatigue, understanding where on the body a sensor could be located, what data various locations provided, what physiological and biomechanical signals could be used and how to model fatigue, how fatigue was measured in laboratory settings, how this could translate to the field, what protocols could be used to validate and calibrate a field environment and finally what models could be used.

Significance of Contribution

Fatigue is the result of a complex interaction between both actual and perceived effort guided by the assessment of available resources which is informed by prior training, immediate motivation, perceptions of environment, capability and goals. Fatigue has been extensively studied in the laboratory for cognition on pilots, workers, patients and drivers and for physical fatigue on athletes. However multiday missions in extreme environments require both cognitive and physical fatigue to be studied in a long duration field environment without affecting the mission goals by distractions or additional encumbrances by assessments or sensors as low compliance will result.

This research addressed the four critical gaps in the current literature (i) how to a-priori assess compliance to a protocol in the field, (ii) how to calibrate a field protocol with labels for machine learning to enable a comparison to laboratory equivalents, (iii) how to determine if terrain modulation could be included in a dataset to achieve laboratory equivalent accuracy, and (iv) how to model fatigue with a minimum of sensors and user interaction in a field based environment.

The studies in the thesis have resulted in a fieldable long duration protocol and data pipeline with validation to existing assessments for physical and cognitive fatigue. The protocol tested aspects of challenges in the field including: (i) data from limited sensor options for high participant compliance, (ii) no direct observation of participants and (iii) terrain and time of day/night gait modulation.

The research firstly addressed possible sensors and protocols by determining a novel equation for compliance that considered goals, assessment work and motivation. Secondly a field protocol and calibration was researched and compared to laboratory human activity detection. Thirdly the field protocol was used to validate an automatic fatigue model compared to existing standard assessments.

This thesis has contributed new tools and protocols to the field by enabling long duration field research into fatigue and performance that is scalable with high levels of participant compliance, low sensor count and manageable data analytics, to achieve a prediction model for an individual. The protocol enabled comparatively high frequency assessments which was required for machine modelling and to understand short term fatigue.

The research validated the protocol compared to previously published laboratory results with an equivalent amount of data. The research added real world noise sources from terrain variation with an equivalent number of data. The 11 million data points from the research was comparable to previously published laboratory datasets and took six months to label which is a known problem in machine learning.

Thesis Overview

This thesis (Figure 1-3) is a combination of thematic sections and papers that have been submitted or published in various journals. Table 1-1 shows research key points from each chapter and the links between each chapter with the themes and research conducted.

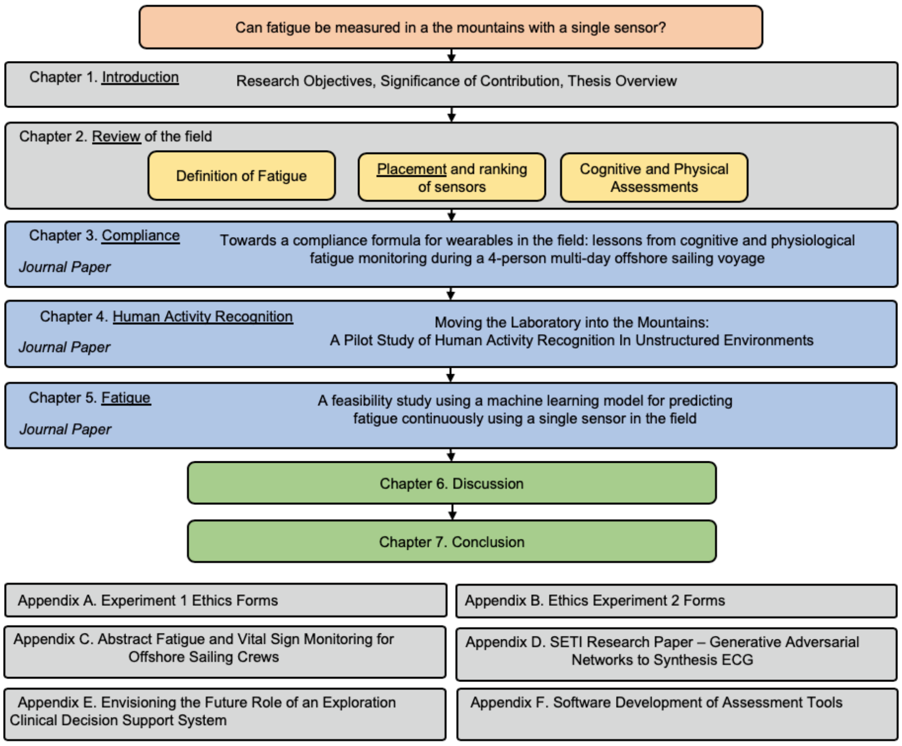


Figure 1-3 Thesis structure

The first thematic section of the thesis (Chapter **2**) is a narrative literature review covering the subject of fatigue with respect to the fields of physiology, psychology and machine learning. The chapter then discusses placement of sensors with the associated signals that are possible from locations on the body and assessments to determine cognitive and physical fatigue. The first field study for this thesis investigated fatigue protocols in extreme environments during a 12-day off-shore sailing voyage from the east coast of the United States to Antigua in the Caribbean with four crew members. A protocol of assessments for cognitive and physical fatigue was established, ethics preparation and approval was granted and the data collection took place. The outcome of this research was useful to understand protocol compliance in long duration field activities. There were significant learnings on the requirements for software assessments, data analysis labelling and compliance. This outcome lead to a side study of compliance and a formula put forward to allow a trade-off analysis by predicting compliance to a protocol based on: goals, task work-loads and motivational factors. This work resulted in **Chapter 3:** Russell B.K., Hume P.A., McDaid A., Toscano, W., Towards a compliance formula for wearables in the field: lessons from cognitive and physiological fatigue monitoring during a 4-person multi-day offshore sailing voyage, Journal of Human Performance in Extreme Environments (submitted).

With the lessons from the first field trial and the insight from the compliance formula it was decided that custom software was required in order to achieve adequate compliance for fatigued participants in the field. Given a year was spent on collecting and evaluating the data, it was also decided to develop a protocol and trial it on a single participant to determine its efficacy and inform requirements for further work.

The outside trail run used for the research was calibrated with terrain slope and surface to initially determine if a deep learning model could be used to classify human activity recognition (HAR) with sleep deprivation and fatigue as the independent variable with the additional challenge of no observer to label activity. A key finding from this work was the significant amount of time to label high frequency accelerometery data in order to start training AI models. This HAR experiment proved successful and attained similar accuracy results to equivalent results published for laboratory experiments using predictable and smooth surfaces, no obstacles and a researcher observing to label data. The protocol to calibrate the track and data pipeline for HAR classification is described in **Chapter 4:** Russell B.K., Hume P.A., McDaid A., Toscano, W., Moving the Laboratory into the Mountains: A Pilot Study of Human Activity Recognition In Unstructured Environments, IEEE Transactions on Biomedical Engineering (submitted).

The final aim of this research was to establish the efficacy of predicting cognitive and physical fatigue in the field using a single sensor for maximum compliance. The protocol was designed with intentional restrictions to ensure the protocol is scalable in the future and applicable to translational research in the field. **Chapter 5** describes the various assessments used and which were sensitive to the fatigue protocol, the data pipeline for a regression model to predict fatigue, and the results. This work resulted in **Chapter 5**: Russell B.K., Hume P.A., McDaid A., Toscano, W., Predicting fatigue with a single sensor and machine learning in mountain terrain, Journal of Neurocomputing (submitted).

A discussion of the overall thesis work is presented in **Chapter 6** with a summary of the findings discussed, along with the limitations and potential future directions possible.

Table 1-1 Research key points from each chapter and the links between each chapter with the themes and research conducted

Multi Day Fatigue Computation using Artificial Intelligence and a Single Sensor in an Uncontrolled Environment

Outcome:	The thesis investigated if a single wearable sensor with an AI model could accurately predict cognitive and physical fatigue in an unstructured environment with obstacles and terrain variation. A protocol was developed to calibrate an outside trail running track and label activity data from an inertial measurement unit and electrocardiogram. A data pipeline was developed to process the new dataset and train a convolutional neural network and predict fatigue.

Chapter No.	Chapter Title	Chapter Content-Questions/Rationales/Findings
1	Introduction	MAIN QUESTIONS OF THE THESIS: 1. How to get compliance to a protocol in the field? 1. How to conduct research in the field accurately and with validation? 2. Can a protocol be developed with terrain and fatigue as independent variables in the field with laboratory smooth terrain comparable accuracy? 3. Can a single sensor and AI model achieve an accurate prediction of human activity recognition? 4. Can a single sensor and AI model predict fatigue? RATIONALE FOR THE QUESTIONS: The key outcomes of the thesis were to develop: a) A compliance formula to assess various tasks in a protocol to predict compliance a-priori b) A protocol for field research c) A cognitive assessment iPhone application d) A python graphical tool for cognitive load (MATB) e) A data pipeline and model to classify activity from field data f) A data pipeline and model to calculate fatigue from field data
2	Review of the Field	QUESTIONS: 1. What is the definition of fatigue? 2. What sensor data is available from various positions on the human body?

		3. What cognitive and physical assessments can be used to validate filed work? APPROACH Literature review FINDINGS • Fatigue is a result of complex overlapping systems. • Fatigue is a multi-disciplinary topic including physiology, neurology and psychology. • The definition of fatigue requires context from the field • A subset of available sensors are possible when considered against the limitations imposed by remote field environments and exercising participants. • There are various cognitive and physical assessments possible that can be applied to field environments which cover different types of cognitive performance deficits. NOVEL CONTRIBUTION • Recommendations on a definition of cognitive and physical fatigue for remote field environments with recommended sensors and assessments. • Gaps in literature identified for field-based research of fatigue.

⇕	Link between Chapters 2 & 3:	Having identified the definition, sensing and validation mechanisms for cognitive and physical fatigue the next step was to perform an field study to assess the practicality of the protocol and sensitive of the assessments.

3	Compliance for Wearables	QUESTION: • What is the sensitivity of various assessments? • What protocol makes sense in the field? • How can compliance be calculated ahead of time to design a protocol? RATIONALE FOR THE QUESTION: • To assess the validity of the approach in a real-world environment. APPROACH • A 12-day 4-person offshore sailing voyage was used as the fatiguing field environment. FINDINGS • Independent applications do not work successfully when used multiple times on multiple people. A custom application is required for assessments.

		• Cognitive assessments are sensitive to various loads and stimuli. • Compliance can be low in a fatiguing and challenging environment. • A compliance calculation is possible which can give insight into how a protocol can be designed. NOVEL CONTRIBUTION • A compliance formula was put forward to calculate compliance to an activity.

⇕ Link between Chapters 3& 4: Having developed a set of software tools and a data pipeline it was time to perform a field study to validate the approach.

4	Moving the laboratory into the field	QUESTION: • Can Human Activity Recognition (HAR) be performed in a remote environment and compared to laboratory grade assessments. RATIONALE FOR THE QUESTION: • There are many trade-offs and decisions to be made when moving from the laboratory to the field. The protocol needs to be implemented in a real world environment to assess it practicality. APPROACH: • Develop a protocol to address the various challenges in data collection in a remote environment with no direct observer. FINDINGS: • A protocol with wearables sensors and assessments of fatigue, both cognitive and physical, can be performed in a remote environment for a long period of time. • A periodic protocol allows laboratory grade assessments for validation. • Data wrangling to align data over time and geography can be automated but requires a large amount of manual labour when dealing with over 3 million data points for multiple data types. • Terrain variation and obstacles from a real-world environment generate significant variation in sensor readings. • It is possible to train an AI model to accurately classify HAR. • AI and a single sensor can accurately classify human activity. NOVEL CONTRIBUTION: • A field protocol to calibrate an outside trail running course over terrain and obstacles. Validation that HAR can accurately be assessed in a remote field environment in the presence of terrain variation, obstacles, fatigue and voluntary activity.

⇕	Link between Chapters 4 & 5:	Having shown the protocol worked for human activity recognition the next step was to determine if any of the assessments were sensitive to the protocol and train a regression model for fatigue.

5	Machine learning model for fatigue prediction in the field with a single sensor	QUESTION: • Can fatigue be predicted in the field using a single sensor and AI model? RATIONALE FOR THE QUESTION: • Assessments are possible for research but not in real world environments. APPROACH: • Design a protocol for high compliance in the field which enables dataset labelling for machine learning. Furthermore, use the protocol to fatigue a participant to failure while collecting valid assessments. Train an AI model using wearable sensor data against the assessments and determine if they can accurately predict fatigue. FINDINGS: • It is possible to predict physical and cognitive fatigue using a single wearable sensor and AI model in a remote field environment with no observer. • Cognitive fatigue significantly increases perceived exertion and physical loads should be reduced when used in combination with cognitive loading. NOVEL CONTRIBUTION: • A fieldable protocol, software toolset and data pipeline for remote environments using a CNN regression model to accurately predict cognitive and physical fatigue in the presence of terrain variation and obstacles and voluntary activity.

6	Discussion	CONCLUSION: • Fatigue is a complexity of redundant overlapping systems and a definition requires context to be applicable. • Compliance is calculatable and needs to be designed into the protocol to be successful in the field. • A single sensor when coupled with a deep learning model can accurately classify human activity recognition.

		• Fatigue can be predicted in the field using a single sensor and trained AI model. • Further work is required to automate the data pipeline and protocol to enable multiple subjects to be studied in the field. • The inter person variation and generalizability is the next step for this study.

Appendix F	Equipment and Software Development	QUESTION: • What is the best platform to deliver assessments and cognitive loads in the field and how should they be implemented? RATIONALE FOR THE QUESTION: • The device and software should be fieldable in remote environments and used best practice from the literature. APPROACH: • Apple research kit was used as a library of cognitive and physical assessments. The application was designed and a third-party app developer wrote the code using the libraries from the Apple Research Kit. • The Multi Attribute Test Battery (MATB) was identified as an appropriate cognitive load and the thesis author implemented a version in Python. • A data pipeline was developed in Python to process and analyze disparate datasets to align with time, geography and sample rate. NOVEL CONTRIBUTION: • A set of software tools to gather data on assessments, load and data wrangling in a high compliance methodology.

Research Publications Resulting From This Doctoral Thesis

Chapter 3. Russell B.K., Hume P.A., McDaid A., Toscano, W., Towards a compliance formula for wearables in the field: lessons from cognitive and physiological fatigue monitoring during a 4-person multi-day offshore sailing voyage, Journal of Human Performance in Extreme Environments (submitted).

Chapter 4. Russell B.K., Hume P.A., McDaid A., Toscano, W., Moving the Laboratory into the Mountains: A Pilot Study of Human Activity Recognition In Unstructured Environments, IEEE Transactions on Biomedical Engineering (submitted).

Chapter 5. Russell B.K., Hume P.A., McDaid A., Toscano, W., A feasibility study using a machine learning model for predicting fatigue continuously using a single sensor in the field, Journal of Neurocomputing (submitted).

2. REVIEW OF THE FIELD

This chapter is a detailed review of the definition of fatigue and the contributing mechanisms which overlap in a redundant way to allow optimal human performance. The discussion of fatigue covers views from exercise physiology, neuroscience, psychology and machine learning. This section also describes the wearable sensors and the signals available from various parts of the body. Cognitive and physical assessments are reviewed. This chapter concludes with a recommendation of sensors and validation assessments that can be used to quantify fatigue in a field environment.

The most influential factors on a person performing a task safely and competently is their physical ability and their mental ability to interpret a situation, plan and act. This has been explained by Boyd[1] with the OODA loop as shown in Figure 2-1.

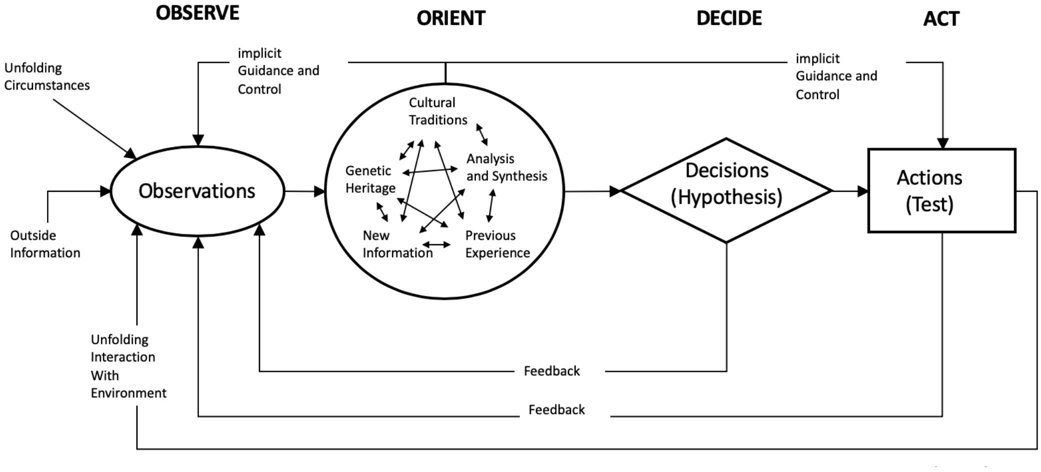


Figure 2-1 John Boyd's OODA Loop Diagram

Physical fatigue and cognitive fatigue limits a person's ability to observe, orient, decide and act hence reduces their capacity to attain a mission goal. This reduction influences acute reactions to events[2], strategic decision making[1], avoiding errors[3] and physical performance such as gait[4–10].

This thesis is studying fatigue in multi day missions where high levels of self-paced physical exercise, strategic thinking, environment cognisance and rapid reaction to unexpected events are the requirements of a person to succeed.

Definition of Fatigue

The study of fatigue is a multidisciplined task encompassing neurology, psychology, physiology, biomechanics and mathematical modelling. Fatigue definitions have been strongly argued from the perspective of cardiorespiratory versus musculoenergetic with recent work including complexity theory and psychology. A timeline of this work is shown in Figure 2-2. Fatigue research over the last century have been a case study in scientific discovery were simplified experiments in the lab did not consistently give the answers predicted on theory and new discoveries have increased the models complexity in order to attain field results that align with those theories. Fatigue studies started with chemical and gas exchange models, added neurology then cognition and now includes the brain almost as though it is running software to decide how it feels and how much it wants to perform.

Fatigue can be defined as the capability to perform a task. This combines cognitive and physical performance. A practical definition of fatigue requires the context of the goal or mission, as the environment and performance tasks requiring a mixture of physical and cognitive tasks defines the type of fatigue to be considered. For driving a car this has been defined as the “time to react to an unexpected event” [2] including visual, cognitive, 3-dimensional modelling, prediction and neuromuscular responses.

The definition of fatigue can be viewed from the resulting performance in a given situation usually due to a cascade of reduced resources in the bodies overlapping redundant systems[2].

Physical Fatigue

Figure 2-2 summarises the historical approaches to understanding fatigue. There is a description of the underlying theories used to define fatigue in the context of this thesis. The inclusion of several older papers gives context to the development of the concept of fatigue and indicates that this field may remain an area of contested hypothesis as the models continue to get more complex and require additional fields of expertise. Recent papers by Noakes[11] and his collaborators have shown a healthy inclusion of multiple disciplines and various thinking. Complexity of systems and large datasets are currently problems where artificial intelligence models are proving successful.

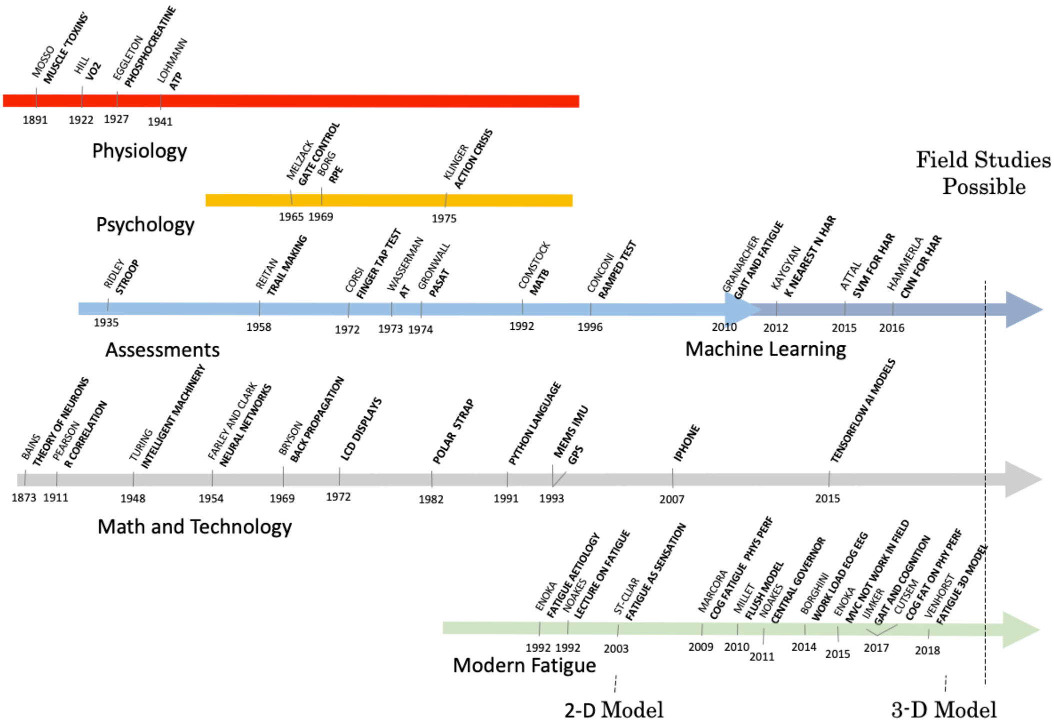


Figure 2-2 Timeline of major scientific contribution leading to fatigue research

Exercise Work Load Based Fatigue

Physical fatigue in athletic situations has been discussed for over 100 years. The current theory on exercise based fatigue starts with the various effects of peripheral or central limitations. Peripheral fatigue refers to skeletal muscle with reduced ability to produce or reproduce force. Mosso[12] in 1891 was the first to measure muscular work and show muscle contractility in humans. He was also the first to determine that muscles generated a 'toxin' which limited work, later defined as lactic acid. He then developed a model of exhaustion, pointing out there was two effects on fatigue: the central and peripheral. He also determined that fatigue of the mind reduces the strength of the muscles. He has also reported that increased fatigue can change the 'mood' of a person which speaks to models developed a hundred years later such as the Central Governor Theory[11], Flush model[13] and work by Macora[14] . Hale[15] describes in detail the history of developments in sports and exercise physiology up to 2008, the time of publishing. In 1922 Hill[16] won the Nobel prize for his work on skeletal muscle and maximum oxygen update. He discovered contracting muscle had two phases which generated different temperatures showing aerobic and anaerobic activities. In the same year he presented a paper on the “oxygen consumption during running” where he was the subject and measured oxygen uptake with an indication of a maximal uptake volume. Hill collected data to show sub maximal exercise could be maintained over time and greater than maximal exercise resulted in cessation. Phosphocreatine was first recognised in 1928 by Philip and Grace Eggleton and adenosine triphosphate (ATP) was discovered in 1941 by Lohmann but not fully modelled in the muscle unit 1962 by Cain and Davis.

Early efforts to quantify fatigue were based on a threshold concept where a maximum rate of consumption or generation had been reached and was unsustainable over time. The anaerobic threshold (AT) is defined as the maximal work intensity where blood lactate increases rapidly leading to metabolic acidosis and exercise cessation. Protocols were developed to determine the value that ended in cessation of an activity at peak power or speed. Various non-invasive tests have been validated to determine AT. Wasserman and McIlroy[17] in 1973 showed that AT can be detected non-invasively by monitoring respired O_2 and CO_2 gases. Further work showed that the threshold occurrence was not steady state and required a buffer of bicarbonate to be consumed before respired gas ratios were affected. Conconi[18,19] in 1988 reported a method to determine AT by measuring the deviation of a linear relationship between speed and heart rate. Both of these measures have been refuted and argued yet both methods are still in use due to their non-invasive nature.

In hindsight it is interesting to see how maximal test results varied and could not be explained without today’s more complex definitions of fatigue. Often the protocol was designed to achieve repeatable results, but this was inadvertently hiding the underlining mechanisms that take place in an uncontrolled field environment. The field moved into an area of thought where central regulation could start to take a role.

In 1965 Melzack[20,21] published the gate control theory of pain. This introduced the concept that the brain’s perception of pain may be modulated by cognitive activities and is not strictly related to the amount of signal in nerve fibres. This is the first indication in the literature that the brain was not reacting in a linear way to physical signals and paves the way to an explanation of the variance in MVC experiments and pacing strategies in submaximal activities.

In 1975 Klinger[22,23] introduced the concept of action-crisis where a person's motivation to attain a goal is modulated by the level of effort or complexity to attain that goal. The persons motivation to reach a goal decreases incrementally such that when the goal is dropped the persons mental and physical state is protected. This work was to be picked up decades later when motivation was considered in the performance fatigue paradigm.

In 1982 Borg[24] published his rating of perceived exertion (RPE), a gestalt was proposed as an integration of the body's signals from peripheral joints and muscles and the cardiorespiratory system.

In 1992 Enoka[25], introduced the concept that fatigue is task dependant. Additionally, he specified that the effort to perform a task can be independent of that actual force generated. This introduces the concept that fatigue may be limited by perception of exertion or actual physiological exertion.

In a 1997 lecture and later in 2001 Noakes[11] et al introduces three models of fatigue: peripheral, central and cognitive. The peripheral model describes how muscle metabolite concentrations control exercise intensity. This model includes limiting factors such as metabolic accumulation in active muscles, insufficient oxygen delivered by the cardiorespiratory system and neurotransmitter depletion in the brain's motor cortex reducing the drive to the muscles. The central fatigue model is seen as a subconscious process limiting exercise to maintain the system within safe limits and safe from damage. This system may reduce intensity or cause the cessation of activity. In this model maximal peripheral effort is modulated centrally and never reached. Previous studies have shown fatigue may occur upstream of the motor cortex in other cognitive processes. The cognitive model considers fatigue as a continuous conscious state that regulates intensity using afferent feedback with constant comparison to prior experience to modulate intensity. Noakes presented a model where fatigue is a sensation or emotion. He points to chronic fatigue syndrome as evidence of individuals that are capable of muscular force comparable to other sedentary individuals. Noakes continued the argument that fatigue may be located in a specific region of the brain which has been controversial.

In 2001 Noakes et al posited the central governor model (CGM) where a..

> *central neural governor determines cardiac output by regulating the mass of skeletal muscle that can be activated during maximal in both acute and chronic hypoxia.*

> *Noakes saw differences between exercise cessation before all peripheral muscle fibres were activated, indicating an overarching control mechanism. The CGM allows the brain to maintain homeostasis of the peripheral systems by anticipating the load and available energy. He also notes that previous protocols have controlled the pace of the athlete and this hides the effect, whereas the only way to determine the full system is to allow self-pacing.*

In 2003 St Claire-Gibson and Noakes[26] put forward a model that was more complex than the traditional models of fatigue that were based on peripheral or central limitations. The model posited that the brain causes termination to avoid catastrophic failure of any peripheral system.

In 2009 Marcora[14] reported that cognitive fatigue impairs physical performance where the level of cardiorespiratory and musculoenergetic factors stayed relatively constant. Self-reported success and intrinsic motivation were also constant. Cognitively fatigued subjects reported higher RPE levels and ceased exercise earlier.

Marcora[14] introduced a model of a psychobiological state where mental fatigue can be the limiting factor to performance by heightening perception of effort rather than cardiorespiratory and musculoenergetic processors. This introduction of mental fatigue limiting physical performance is an important addition to the previous physiological limiting mechanisms where cognition was not treated as a limited resource.

In 2011 Noakes summarised why a centrally controlled model is required to explain certain observed end states:

> *(i) ... differential pacing strategies for different exercise durations; (ii) the end spurt; (iii) the presence of fatigue even though homeostasis is*

> *maintained; (iv) fewer than 100% of the muscle fibres have been recruited in the exercising limbs; (v) the evidence that a range of interventions that act exclusively on the brain can modify exercise performance; and (vi) the finding that the rating of perceived exertion is a function of the relative exercise duration rather than the exercise intensity.*

However, there are examples of performance athletes performing to levels where peripheral fatigue was clearly the limiting factor not central fatigue. An example was when Hyvon Ngetich in 2015 crawled the last 50m of a marathon.

In 2011 Millet[13] discussed how maximum voluntary contractions (MVC) did not correspond to performance at submaximal long duration exercise including pacing strategies. Millet introduced the Flush model …

> *This model has the following four components: (i) the ball-cock (or buoy), which can be compared with the rate of perceived exertion, and can increase or decrease based on (ii) the filling rate and (iii) the water evacuated through the waste pipe, and (iv) a security reserve that allows the subject to prevent physiological damage.*

This model suggests that central regulation uses afferent signals from the periphery and organs with the addition of signal from peripheral fatigue and spinal/supraspinal inhibition. Millet explained that this model explains the influence of sleep, mental fatigue, pain killers, psychostimulants and the effect of nutritional and cognitive strategies that have been found to affect ultra-marathon performance.

The concept of critical power is a useful model to bridge the work that describes increasing levels of effort to failure and self-pacing. Vanhatalo[27] in 2011 described the concept of critical power as where steady state power transitions to non-steady state.

Enoka proposed that the only way to determine field performance was to measure fatigue in the field, which is the question for this thesis.

Enoka[28] described the work of 18th century Mosso's[12,29,30] two dimensional model which treats performance fatigability and perceived fatigability separately. Enoka concludes experiments that attempt to disentangle causes of performance decrements such as maximum voluntary contractions end up becoming task dependant and the laboratory findings do not translate into real world performance reduction. Enoka proposes …

> *fatigue be defined as a symptom in which physical and cognitive function is limited by interactions between performance fatigability and perceived fatigability.*

In 2018 Venhorst, Micklewright and Noakes[31] suggested a three dimensional model (bio-psycho-social) that includes perceived strain (physical and mental), core affect (valence and arousal) and mindset (flow state and action crisis). They suggested 6 different self-report tests to assess the six inputs to the model; 15 point (6-20) Borg scale, 15 point (0-14) Borg scale, Empirical Valence State scale (EVS), Felt Arousal State scale (FAS), Flow State Scale (FSS) and Action Crises Scale (ACRISS). It is interesting to note that Noakes points out the limitations of a protocol setting an athletes pace which may obscure any central governor effects but offers 6 self-report scales with multiple questions which in itself is a significant cognitive load. The use of questionnaires does not lend itself to field work where self-pacing is required for performance and fatigue assessment. Additionally the time between questionnaires will likely lead to low time resolution which can miss micro recovery. It is my view the field is struggling to understand how to non-invasively measure fatigue for self-pacing in the field.

Exercise physiology has spent a lot of energy on whether fatigue is limited by peripheral (muscular energetic, cardiorespiratory), central (motor drive regulation for homeostasis including) or social aspects (anticipation of needs, goal, motivation, experience, competition). It is fascinating to look back on Mosso's work from 1897 and see indication of the next century of scientific discovery and refinement. In my view cessation of exercise or self-pacing strategies each have examples that show all the models are correct some of the time. Noakes has generated a large amount of commentary from his CGM yet there are examples where central control do not limit peripheral function to the point of collapse, examples include marathoner Sian Welch and Wendy Ingraham crawling across the finish line in the 1997 ironman after collapsing multiple times in the last kilometre. I would suggest all models are correct some of the time given the cause for self-pacing and cessation varies between cases.

Fatigue in any one instance is the failure of the first cascade running out of resources. Each instance may be due to a different failure mechanism so all models are valid, but a different one may dominate each time.

The aetiology of fatigue is the very reason why a systems approach is required vs a reductionist approach. Fatigue experiments need to be in the field given the environment and goals largely contribute to fatigue.

If muscular temperature, lactic acid or lack of motivation from action-crisis first limits performance then that model explains the current situation. However, a change in environment, nutrition, hydration or circumstances may rapidly and dramatically change performance until another reason becomes the limiting factor.

Cognitive Fatigue

Cognitive fatigue can be defined in a similar way to physical fatigue in that there is a difference between perceived effect and actual performance. Wylie[32] reported traumatic brain injury (TBI) patients have a high perceived mental effort while performing at a similar level. Mental fatigue can be defined by performance requirements for a particular task. Cognitive fatigue can be viewed as a combination of goal, adaption and reward trade-offs including the energetic requirements to achieve a goal[33,34].

Borghini[35] in 2014 has shown that increased EEG power in the theta bands and decreased energy in alpha bands occurred at high mental work load. Successively increased power in theta, alpha and delta bands occurred at the transition from mental workload to mental fatigue. An increase in mental workload reduced situational awareness thus reducing operator performance. Borghini stated:

> *.. mental fatigue is believed to be a gradual and cumulative process and is thought to be associated with a disinclination for any effort, a general sensation of weariness, feelings of inhibition and impaired mental performance, reduced efficiency and alertness.*

This thesis is focused on quantifying the degree of cognitive fatigue at a given time by the degradation of performance as it interacts with the environment, goals, physical and cognitive performance. Cognition can be viewed by different axes; assessments, cognitive domains or brain regions recruited. Cognitive fatigue can be defined as the reduction in relevant cognitive performance to the mission.

Cognitive abilities, performance and fatigue are all related. Operational psychology of performance involves various specialities including clinical, forensic, social and industrial psychology. For this reason a brief history of the fields development is not as succinct or linear as the physiology historical equivalents.

Performance was aptly defined by Aoyagi and Stevens[36]:

> *We define performance as a process of developing one's knowledge, skills, and abilities (KSAs) in a given performance domain and then recalling and demonstrating these KSAs during a discrete performance event*

Based on this definition we can use a job task analysis (JTA) to determine the cognitive domains required for a game, goal or mission and use objective testing to determine the use of the persons KSAs to an appropriate battery of tasks.

Performance psychology is a further refinement on the Aoyagi-Stevens[37] definition to focus on the understanding of psychological factors for superior performance under stressful conditions. Aidman[38] introduced the concept:

> *Cognitive Fitness (CF) as a multifaceted and differentially malleable capacity to deploy cognitive resources, knowledge and skills to meet demands of operational task performance throughout a career"*

Cognitive fitness covers optimising KSAs by looking at preparation, in-field performance and recovery.

Cognitive fatigue for multi-day missions can include similar tasks from several applications of industrial, sports and military to asses which cognitive performance and fatigue criteria should be considered. Cognitive fatigue has been studied extensively in aviation[39–44], industrial[45,46], driving[2,47–50] and Military[51] operational situations. Cognitive domains need to be listed that are required to successfully carry out a mission. NASA recently developed an assessment tool called Cognition[52] to include cognitive domains; abstract reasoning, attention, computation, emotion processing, risk taking, spatial orientation, spatial processing, working memory and visual spatial working memory.

Fatigue Summary - Context Specific Definition of Fatigue

The current view of fatigue models is that peripheral, central and cognitive fatigue all interrelate, while different scenarios or protocols may focus on a failure mechanism or strategy for self-pacing.

Every situation and mission will have different load factors, environments, complexity and challenges. Every participant will have slightly different level of preparation in each mission or competition. A method is required that can determine fatigue as the sum of performance.

Optimal performance requires many systems to be actively working towards the required level of effort. Cessation of effort occurs when a particular cascade of redundant resources is depleted or is estimated to be depleted for the given goal at the current state of work. If the central model is out of capacity from lack of sleep, dehydration, heat or cognitive overload it may be the cause of reduction. If conscious distraction or stress occurs then peak levels of effort may not be possible. Fatigue is a ArgMin[53] function of a complex system. In a mission reduced performance will vary within each of these tasks. The central governor theory introduced concepts and language that enabled new ways to think about the field of fatigue even though there is debate in whether the governor is a local or whole brain response.

Table 2-1 History of exercise fatigue with key contributions to the field

Year	Author	Key Contribution to Fatigue
1891	Mosso[12]	Muscle workload and muscle fatigue including the Ergograph to graph fatiguing muscle repetitions. First to show 'toxins' in muscle causes fatigue not just lack of oxygen.
1922	Hill[16]	Aerobic and non-aerobic muscle metabolism
1927	P & G Eggleton[54]	Phosphocreatine
1941	Lohmann[55]	ATP, Adenosine Triphosphate
1965	Melzack[20,21]	Gate control theory
1969	Borg[24]	Borg scale Rating of Perceived Exertion, RPE
1973	Wasserman[17]	Anaerobic Threshold
1975	Klinger[22,23]	Action crisis
1988	Conconi[19]	Ramped effort test for anaerobic threshold
1992	Enoka [25]	Fatigue aetiology
2003	St Clair-Gibson [26]	Fatigue as a perceived sensation
2009	Marcora [14]	Mental fatigue impairs physical performance
2011	Noakes[11]	Central Governor Model
2011	Millet[13]	Flush Model – perceived fatigue vs performance fatigue
2014	Borghini[35]	EEG increases with mental work load
2016	Enoka[28]	MVC vs submaximal do not corelate Language and experiment design are lacking to understand fatigue in the field Psychophysiological = homeostasis + motivation
2018	Venhorst, Mickleworst, Noakes[31]	3 way model = bio-psycho-social

Data Available That Correlates With Fatigue

Previous fatigue studies have used computer interaction[56], accelerometery, electroencephalogram (EEG), Electrooculography (EOG)[57], electromyography (EMG) and electrocardiograph (ECG)[4,58–60]. However, this measurement equipment can be difficult to use in the field. In 2017 Aryal[61] used skin temperature, heart rate (HR) and EEG to determine fatigue in construction workers. Borghini[35] in 2014 outlined measures of fatigue and mental work load in drivers and pilots using EEG, EOG and HR.

In 2014 Zhang[62] used EEG, EMG and EOG with artificial neural network (ANN) to determine fatigue where the subjects were sitting and performing a switch task on a computer screen. Parkinson's disease is interesting as a possible proxy for cognitive fatigue. Kirchner[63] in 2014 used shoe sensors on Parkinson's patients to measure stride time, he also used detrended fluctuation analysis and entropy to analyse the data. The results showed healthy subjects had low variance in the gait and Parkinson’s patients approached a more random variation.

The relationship of sensor data and reduced cognitive and neuromuscular function have been studied in various populations from healthy individuals in long term sports, occupational safety studies and various disease states which affect cognitive and neuromuscular systems.

Ocular measurements have a rich dataset[35,64] which correlates with many physiological states and has been used in cognition, spatial planning and fatigue. Dinges[65] has reviewed various technologies using ocular measurements against the psychomotor vigilance test (PVT). The Volkswagen[2] motor company has a Driver State Monitoring System [2], which uses blink, head nodding and eye pupil position. This system worked in a research vehicle with the driver in a fixed position and it occasionally suffered from intermittent occlusion of the drivers face. Ocular Electromyography (OMG), could be used if a suitable sensor attachment could be found or built.

ECG derived features such as heart rate (HR), heart rate variability, (HRV), and heart rate complexity (HRC) are all potentially useful for a fatigue model. HR was used as a direct measure of physical load and provides a measure of core temperature through increased cardiac output due to increased volume demand from vasodilation. Additionally, core temperature was modulated by circadian rhythms which modulate the participants’ cognitive system for a multi-day event. HRV provides a measure of the autonomic system modulation by both physiological load and central control via hormones such as adrenaline. Guidi[66] used HRV for the detection on muscle fatigue with a 78% accuracy showing a relationship between muscle fatigue and HRV. HRC[67] has been shown to corelate to mental stress and cardiac health so will also be used for the model.

Cognition and gait has been shown to be linked with aging[68,8], fibromyalgia[69], Parkinson’s[63] as well as multiday athletic events with reduced sleep[70]. Multiday events modulate affects with changes in sleep patterns and circadian rhythms[71]. Fimm[72] discussed the relationship of visio-spatial attention and arousal which was important as spatial planning was a key part of most multi-day events. Granacher[73] showed that tasks affected gait characteristics. Fuller[74] has shown gait parameters changing in runners that are overreaching during running. In conclusion, accelerometery was a useful measure of all the systems (vestibular, ocular, goal setting, cognitive spatial planning and neuromuscular) in 3D space. However, human activity recognition[75] may be required to allow intelligent detection of transient gait phases such as lateral sway contextualised to different activities such as walking in unobserved uncontrolled conditions.

Table 2-2 summarises the various measured data parameters affected by fatigue. This is important when deciding the dataset for a machine learning model as the data should have previously been shown to be correlated with mental and physical fatigue.

A determination is required of the sensed data that has shown a sensitivity to central and peripheral fatigue. This will give the neural network model the best chance of finding a solution and also inform the data wrangling required for noise reduction, filtering and feature extraction.

Grobe[68] discussed a connection between ageing and gait changes related to reduced cognitive functioning. Heredia[69] shows spatio-temporal changes in patients with fibromyalgia. This points us in the direction that accelerometer data with feature extraction should be related to cognitive function and hence included in an AI model. Verlinden[8] concluded that "*cognition and gait show a distinct pattern of association*", this is important as gait relates to parameters such as pace and executive function which is important on a multi-day event where sleep and other non-ideal recovery behaviour will reduce executive function and athletic performance[70]. Kirchner[63] applied detrended fluctuation analysis to gait data from Parkinson's patients with cognitive reduction. Van Dongen[71] considered sleep as a homeostatic process which results in reduced psychomotor vigilance through reduced sleep, however, it is modulated by circadian rhythms. Circadian rhythms on multi-day data will appear as seasonal variation within the data and should be able to be detected as a feature and used in any model. Fimm[72] discussed the relationship of visio-spatial attention and arousal which is important as spatial planning is a key part of most multi-day events for tasks such as navigation and obstacle avoidance. Granacher[73] showed that tasks affected gait characteristics. Fuller[74] has shown gait parameters changing in runners that are overreaching during running.

In conclusion, accelerometery is a useful measure of all the systems (vestibular, ocular, goal setting, cognitive spatial planning and neuromuscular) in 3D space, However, human activity recognition[75] may be required to allow intelligent detection of situational data phases such as lateral sway during walking. Ocular measurements have a rich dataset[35,64] which correlates with many physiological states and has been used in cognition, special planning and fatigue.

Table 2-2 Physiological parameters effected by fatigue.

Term	Test	Function	Reference
Balance	Dynamic Balance Control	Cognition	76
EEG	EEG θ ↑ Drowsiness	Drowsiness	77
EEG	EEG ↓ α ↑ sleepiness and visual flow	Sleepiness	77
EEG	EEG↑ β ↑ concentration	Concentration	77
EEG	EEG ↑ β ↑ anxiousness	Anxiousness	77
EEG	EEG ↓ α /θ ↓ alertness	Alertness	77
EEG	EEG α /θ ↑ emotion and ↓ attention	Attention	77
EEG	EEG shift in bands vs fatigue (D+Th+A)/(B+g)	Fatigue	78
EEG	Frontal Theta visiomotor and cog performance	Cognitive Performance	78
EEG	Alpha activation from Occipital to Anterior with fatigue	Fatigue	79
EOG	EOG (Electro Ocular Gram) vertical, horizontal	Attention	78
EOG Ocular	PERCLOS(percentage eye Closure)	Drowsiness	65
Eye	Pupil Dilation (coeruleus-norepinephrine system)	Arousal/Attention (task difficulty, mental effort)	80
Eye	Eye blink (dopamine)	Goal directed behaviour	78
Eye	Scanning randomness	Stress	81
Gait Velocity	Gait Velocity	Cognition	82
HRV	HRV: SDNN, VLF,LF,%HF and LF/HF	Fatigue	83
HRV	HRV	Working Memory	84
HRV	HRV	Time on Task	84
HRV	HRV no proportional to participative measures	Participative measures	84
HRV	SDNN sympathetic and some para-sympathetic	Sympathetic	84
HRV	RMSSD para-sympathetic	Parasympathetic	84
HRV	pNN50 with workload	Work load	84
HRV	HF Fatigue	Fatigue	85
	LF/HF Fatigue (8 hour COG fatigue)	8 Hour Fatigue	85
Gait	Pacing	Mental Fatigue	86,87
Gait	Variability and Pace	General Cognition	8
Gait	Rhythm, timing	General Cognition	63

Resulting Action From Combining Cognitive and Physical Fatigue

Action is the final product of physical activity, mental activity and motivation[88]. While various models give insight into the complexity that results in action, action is the end result that determines performance to win a race or complete a mission without breaking equipment or people. So the measure of fatigue is not necessarily the measure of individual pieces in the action complex as the outcome on action will be indeterminant and time varying.

Subset action tasks that are prior validated can be used to determine action, performance and fatigue. When included in a protocol, these subtasks and assessments can be used to validate a field protocol which attempts to automate the measurements and remove the requirement for subtask assessments to take place.

Psychological modelling by James[89] in 1890 showed how thoughts, feelings and physical interactions leads various brain states.

Atkinson[90] proposed a model for working memory in 1968 shown Figure 2-3. Baddeley[91] in 1974 argued working memory made up of three parts, a central processing part and two slave parts for visual and semantic inputs.

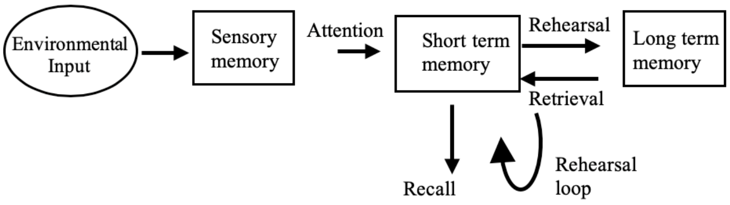


Figure 2-3 Modal model of working memory

Ericsson[92] who in 1993 introduced the concept of 10,000 hours to become an expert at a skill developed his understanding of laboratory to field based research. In his recent 2020 paper he concluded that laboratory work for complex field situations do not always extrapolate and the work needs to be done in the field.

> *There is an emerging consensus that the traditional laboratory approach to studying general learning mechanisms in skill acquisition and extrapolating simple processes, such as rehearsal and strengthening of associations cannot provide sufficient accounts of the acquisition of complex skills and the*

> *complex cognitive processes mediating their gradual acquisition during designed practice activities*[92].

Bier[93] et al in 2020 systematically reviewed monotony based fatigue. Fatigue for driving tasks were assessed. Fatigue was modelled in three distinct models; (i) sleep related fatigue influenced by circadian rhythms or insufficient sleep, (ii) active task related fatigue when a person cognitive resources are over taxed, (iii) passive task related fatigue is due to monotony from underutilization.

Cognitive and Physical Assessments

Neurocognitive testing has traditionally been administered one on one with a clinician. It can take up to several hours and is performed in a low stimulus environment. This approach does not lend itself to in-field research and recent work has been carried out to enable neurological computerised assessment tools[94] (NCAT). NCAT tools offer logistic advantages in time and do not require administration of a clinical expert. Reaction times can be measured and large variation can be introduced to mitigate training effects. It should be noted that NCAT equivalents are not identical and require unique normative data.

There is a large amount of data on neurocognitive performance tests and their applications. A systematic approach is required to determine the cognitive deficit or performance of interest and match the available evaluations based on the circumstances the tests will be applied to[95]. One aspect of choosing an appropriate cognitive assessment is sensitivity to the expected cognitive variations during the protocol.

Assessments typically include multiple individual tests, each testing a subgroup of cognitive domains. Assessments batteries used historically include ANAM[96] (soldiers), CNS-Vital Signs[97] (clinical trials), Cognition[52] (Astronauts) MATB[98] (pilots). When selecting neurocognitive assessments the process typically starts with listing functional requirements from a JTA or using tests that have previously shown sensitivity in similar applications.

For a multi-day mission in the mountains, tasks included physical coordination in uncontrolled environments, strategic planning, navigation, situational awareness, decision making, anticipation of changing events, motivation and central drive, valence and arousal at appropriate and sustainable levels.

A battery of assessments were chosen that have previously shown sensitivity to the protocol loads and fatigue related diseases. These included assessments used for fibromyalgia[99], Parkinson's[100] as well as physical[101] and cognitive fatigue[102]. While clinical uses of these tests often take up to several hours it is common practice to shorten the time burden in the field for research purposes. Assessments used included Stroop, Finger Tap Test (FTT), Trail Making A, Trail Making B, paced serial addition test, (PVSAT) memory, and jump height.

Table 2-3 Interactive tests of fatigue suitable for in-field use.

Term	Test	Function	Reference
ACE-R	Addenbrookes' Cognitive Examination Revised	Cognitive Function	[103]
ANAM	Automated Neuropsychological Assessment Metrics	Cognitive Function	[96]
AVLT	Auditory Verbal Learning Test	Cognitive Fatigue	[104]
BART	Balloon Analogue Risk Task	Risk taking, impulsivity	[52]
BESTest	Balance	Somatosensory processing, spatial planning	[105]
Berg Balance Scale	Berg Balance Scale	Somatosensory processing, spatial planning	[105]
BMDT	Basel Motor Dual Task Test Gait velocity + SM + MMSE	Cognition	[106]
Borg RPE	Rating of Perceived Exertion	Physical Strain	[24]
CDT	Clock Drawing Test	Cognitive impairment, Visuospatial, Dementia	[107,108]
CPT	Connors Performance Test	Attention	
Digit Span Backward		Working Memory	
Digit Symbol	Digit Symbol SAHS vs CFS	Cognitive Fatigue	[109,110]
EQ-5D-5L	European Quality of 5 Dimensions	Health Related Quality of Life assessment.	[111]
FTT	Finger Tapping Test, slower with SAHS vs CFS	Neurocognitive	[112]
Jump Test	Jump test	Neuromuscular Fatigue	[4,113]
KSS	Karolinska sleepiness Scale	Sleepiness	[114,115]
LDST	Letter Digit Substitute Test	Neurological disfunction	[116]
LR	Logical Reasoning	Logical reasoning	[117]
N-BACK		Working memory	[32]
PANAS	Positive and negative Affect Scale		[118]
PASAT	Paced Auditory Sequence Assessment Test	Working Memory	[119,120]
PSQI	Pittsburgh Sleep Quality Index	Sleep Quality	[121]

PVT	Psychomotor Vigilance Task	Cognitive Fatigue	4,96
MFIS	Modified Fatigue Impact Scale	Cognitive Fatigue	122
MATB	Multi Attribute Task Battery	Cognitive Load	40 123
MMSE	Mini Mental State Examination	Cognitive Load, Cognitive impairment	124
MSLT	Multiple Sleep Latency Test	Tiredness, sleep pressure	125
N-BACK	N digits back	Working Memory	126
RT	Reaction Time Task	Acute attentional demand	76
RSME	Rating Scale on Mental Effort	Mental Effort	127
SM	Semantic Memory (enumerating animal names)	Cognition	10
SPS	Samn - Perelli Fatigue Scale	Mental Fatigue	128
SPAN	Span Tasks	Working Memory	129
STROOP	Stroop Test	Cognitive Control	130,131
TLX, NASA	Task Load Index	Mental effort from current task	127
TOL	Tower of London	Executive Function - planning	110
Trail Making A	Trail Making A	Response timing, sequence tracking	
Trail Making B	Trail Making B	Response time, sequence tracking, divided attention	132
VOLT	Visual Object Learning Test	Working Memory	96
VRT	Visual Response Time	Cognitive Fatigue	133
WCST	Wisconsin Card Sorting Test	Executive Control - flexibility	110
WinSCAT	Cognition test for space flight		52,110

The tests listed below were selected to be included in the field study as they (i) have been shown to be relevant in the cognitive domain, (ii) had shown previous sensitivity to fatigue and (iii) could be used in the field:

Finger Tap Test – neuro muscular fatigue[112]
Stroop test – cognitive flexibility and selective attention[130,134]
PVSAT – processing speed, attention, working memory[135]
Trail making A and B – motor and executive impairment[132]
Rating of Perceived Exertion - perceived level of exertion[136]
Vertical Jump – neuromuscular fatigue[137,138]

Stroop test is a common cognitive assessment[139,140] which generates a race condition in the brain between reading a colour written on the screen that may be different to font colours: “the participant is shown a series of words that are displayed in colour, and must select the first letter of the colour’s name”. The Apple Research Kit implementation of the Stroop test has been designed to take account the format of a smart phone. This implementation is similar to experiment C of the Comalli Stroop Test[141]. The Apple Research Kit implementation records accuracy and total test time. Total test time was recorded as the direct score for ten words displayed.

Finger Tap Test measures the speed of a finger tapping on the screen and is a measure of neuromuscular fatigue[101,142]. It has been recognised as a suitable test for several neurological disorders including Parkinson's, Alzheimer's disease and acute stroke[143]. The total number of finger taps in ten seconds was recorded.

Trail Making A and Trail Making B has been used to test executive functioning[144] in dementia patients[99] and to determine the relationship between physical and cognitive fatigue for fibromyalgia patients. A series of numbered circles on the screen are connected by tapping in increasing order. Two tests are used (1,2,3,…) and (1,a,2,b,3,c,4,…). The total test duration was recorded.

PVSAT (paced visual serial addition test), is sensitive to partial and selected sleep deprivation[145]. The test has some limitations with reports of the time protocol inducing some stress on the participant. Every three seconds a new digit is displayed and the participant must add it to the previously shown digit. The PVSAT score was the number of correct answers divided by total test duration.

Short Term Visual Working Memory has been reported to reduce with fatigue[146] and is evaluated using the Apple Research kit version of the Corsi Block-Tapping test[147,148]. The original Corsi test used 9 blocks on a 23 x 28 board, where the examiner tapped blocks in a random order of increasing sequence length. The Corsi block tapping test is commonly used by neuropsychologists to investigate nonverbal short-term memory and has been adapted to computer systems[149]. The smart phone version of this test uses a table of 3x3, 4x4 or 5x5 flowers which highlight in random order with increasing lengths of sequences and are tapped in the same sequence. The score is the percentage of correctly tapped blocks answers out of all five tests.

Jump height is a well validated measure of neuromuscular fatigue induced by exercise and mental loads[150–153]. The jump test was a standard peg board test with a back board (200cm to 250 cm, vertical scale 1cm) of interleaved horizontal black and white. A camera was placed side on to the subject at 2.9 m horizontal distance at a height of 213 cm. The camera was an iPhone SE (Apple Inc, Cupertino) 4k video recording at 30 fps. The subject stood beside the peg board and was instructed to hit the pegs as high as possible with their dominant hand. Height measurements were taken from the video as the highest digit measured against the back board with a resolution of 1 cm.

Likert Tests

Several questionnaires were used in the field experiments based on suggestions from experts on expected results or conditions such as sea sickness or sleepiness:

NASA TLX Text – Task Load Index[4]
Mood scale – stomach upset, diarrhoea, vomiting, dizziness[154]
Motion Sickness Scale – nausea, salivation, pallor, sweating, drowsiness, Temperature[155]
Sleep quality- good to poor 4 step scale[156]

Placement and Ranking of Sensors

In previous sections we have discussed the definition of physical and cognitive fatigue with available data and assessments to quantify fatigue. The biomechanical and physiological data collected in the field typically uses wearable sensor.

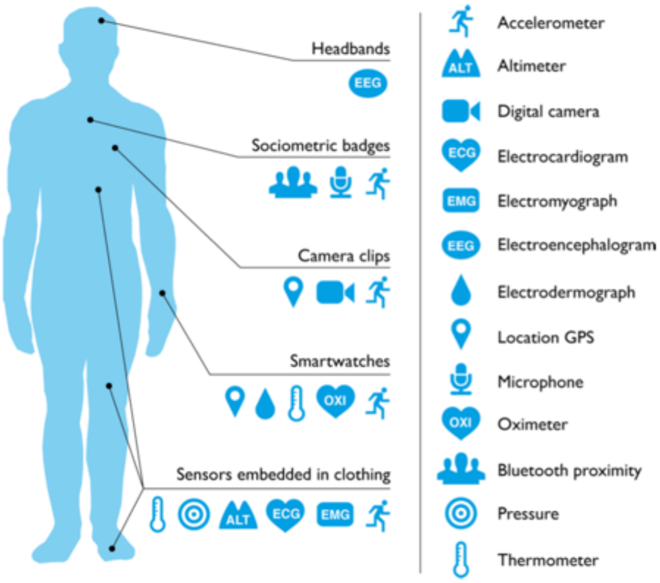


Figure 2-4 Data type available for various parts on the body (Piwek et al 2016, PLOS Medicine)

Figure 2-4 shows viable locations on the body for sensors, this includes temple, wrist, upper arm, ear, posterior neck, sternum, ankle, shoe, or hip. The analysis of wearable sensors for biomechanical state detection using video has been reviewed by Rani & Arumugam[157]. Cell phone sensors have been used by Zhang who identified 32 types of physical activity ($r=0.986$, $p< 0.0001$). However, the device used five sensors on trunk and limbs making it impractical outside the lab[158]. For wearables in the field on endurance personal "wearability" is a major factor to compliance. Sensor data is of no use if the user will not wear the device or keep it correctly fitted. Common wearable positions are head, wrist and sternum, and shank. The wrist is a universal position for everyday use and runners as shown by Benson et al 2018[159], while the most common placement for research is the lower back for trunk and the temple with glasses can be useful for applications where glasses are worn for optical or protection reasons. The temple could give good data for cognitive fatigue if the sensor can be worn. The sternum was chosen as the sensor location for this thesis work given the location gives an estimation of trunk and body movement and also contains the ECG signal.

Table 2-4 Requirements for field-based wearables

Requirement	Value
Battery time	72 Hours or replaceable
Data storage	72 hours
Weight	100 gm
Size	"Small" 30 x 10 10 mm.
IP rating	"waterproof," IP65
Acceleration	+ 6g on body, +20g impact
Perception	does not change gait or behaviour including chaffing or soreness or obstructions during any activity. Cannot be felt to minimise proprioceptive experimental noise.
Signal validity	Can be worn and appropriate vital sign are present during all activity

Artificial Intelligence Applied in the Field

Machine learning is a data-driven modelling technique that learns from data to perform a classification or regression prediction/inference task. The selection of the machine learning model architecture is usually based on application. The model is typically customised from standard model forms. Three systematic reviews have recently been published covering human activity and intensity using imaging and sensors. These reviews provide a comprehensive outline of the relevant literature related to machine learning in the field of human activity recognition. Additional papers are referenced with specific application to HAR and fatigue that were not covered in the systematic reviews.

The first review covers conventional machine learning with popular examples including Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Naïve Bayes (NB), Random Forest (RF), Restricted Boltzmann Machine (RBM) and Artificial Neural Network (ANN). These models require features to be calculated from raw data. Deep learning models have since developed which automatically determine feature morphology, significantly saving analytics time and potentially identifying features not observable to a human. Deep learning models include Recurrent Neural Networks (RNN) and Long Short Term Memory (LSTM) and are often cited as the preferred models for time series data[160] with the advantage of incorporating historical data, however they suffer from the vanishing gradient problem. Convolutional Neural Networks (CNN) are well known for analysing images and have also been used for time series data[161]. They do not suffer from the stability issues of RNNs while enabling parallel processing and requiring less memory which is advantageous when considering wearable deployment. CNN networks have been shown to outperform RNN models[162] and are now considered a good starting point for deep learning models for time series applications. If long term history is required then attention mechanisms can be employed such as temporal convolutional networks (TCN).

The second review, Narayanan[163] et al in 2019 systematically reviewed 53 studies investigating IMUs applied to human activity recognition and intensity. The accuracy ranged from 62% to 99.8 % for states including; sit, stand, lay, walk, stair climbing up/down stairs. Most studies (n=34/52) were conducted in a structured environment, the reason given for this was the complexity of allocating labels in a free living environment. Triaxial acceleration data was used for 90% of the studies. The sample rate was between 10 and 512 Hz. Ground truth data was most commonly collected from direct observation where other methods included photo and self-annotation, which was noted to be unreliable and burdensome. Data processing for activity recognition used both overlapping and non-overlapping windows of length between 1 second to 1 minute. Data for intensity used non overlapping windows exclusively of length from 10 to 60 seconds. The most common models were SVM (n=22), RF (n=19) and ANN (n=19). The paper recommended the researcher should use the maximum number of sensors with the highest sampling rate possible within the limits of battery life time, storage and processing capacity. The review concluded that HAR in free living settings was uncertain and needed further research.

The in the third review, Cust[164] et al in 2019 reviewed machine and deep learning of sport specific movement. This review analysed 52 studies using IMU (n=29) and video (n=22) in sports fields or supervised situations, one study used both IMU and video. 12 of the studies used CNN. Camera utilisation ranged from one to 16 cameras. Data processing included normalisation, outlier adjustment and temporal adjustment. Window widths ranged from 1 to 3.5 seconds. Conventional machine learning using features in order of most used were; SVM (n=16), NB (N=8), KNN (n=8) and RF (n=7). Five studies used IMU's using CNN or LASTM. IMU based studies generally reported classification accuracy greater than 90%. It was noted that supervised learning approaches are 'tedious and time-intensive'. The study concluded experimental set up, data pre-processing and machine learning model development was specific to the sport under study. The study reported that CNN models were a superior approach over conventional machine learning approaches.

Additional attention is worth noting for specific papers. Lockhart et al in 2011 described the generation of the well-known Wireless Sensor Data Mining (WISDM) dataset including accelerometery. They articulate the three design issues for phone-based data mining including; resources limitations (computational, memory and bandwidth), sensor architecture should be scalable to the population, and the results are required real time. A limitation not described in this paper is that the sensor should be worn during the activity and mechanically coupled to the user during high activity levels. Beeck[165] trained four conventional machine learning models with features using a 400m track, 20 participants and six IMU sensors placed in the left and right side of the body fitted to arm, wrist and tibia. This work was valuable as its showed similar results between three different training methods; all runners data, others runners data and individual data. The MAE difference between all runners (n=20) and individual model (n=1) was; arm 0.5%, wrist 13.8%, tibia 2.0%. While this work only investigated runners on a smooth surface it is valuable as the results indicate proving a model with n=1 may be a valid approach for proof of concept research. Lonini[166] in 2016 showed personalised models of HAR can be advantageous as they are trained on less data than global models.

CNN have two key advantages; firstly, local dependency where signals close by are typically related to each other and, secondly, scale invariance where walking and intensity may change but stay recognisable. CNN models have shown good performance on physiological time series data for emotion classification[167], mental fatigue[168] and human activity recognition (HAR)[164,169,170]. Buckley[171] in 2017 used three IMU on runners to classify fatigue. The protocol measured running either side of a fatigue protocol to enable ground truth labelling. Results included 75% accuracy when placed on the lumber spine for the group and 100% accuracy for an individual when placed in the right shank, which is not practical in long duration mountain events where sensors can be immersed under water or caught by vegetation and rocks. Davis[172] did successfully measure runners on road surfaces during the Boston marathon, using sensors on the shank to determine impact forces. Zeng[173] et al in 2014 showed CNN outperformed conventional machine learning models on 3 HAR datasets: Skoda[174] checking a motor vehicle, Opportunity[175] kitchen tasks and Actitracker[176]. Cognitive models have been used for wakefulness detection with accelerometery and ECG[177] and fatigue estimation by Gordienko[178] et al showed positive results with a repetitive exercises in the gym.

In general, most work to date has used conventional machine learning models with increasing use of deep learning with CNN models dominating recent literature due to their computational efficiency and not suffering from stability or excessive memory usage of LSTMs. Advanced deep learning models with newer activation functions and topologies are not currently being used which may be related to how new they are or that simpler models may be more generalisable in the field and easier to implement with restricted computational power. Both systematic reviews give good insight into data processing and sensor data rates for classification and regression which is required for activity identification and fatigue calculations. Narayana's systematic review of activity type and intensity using IMUs gives a good insight into data sample rate, processing and models. Intensity is related to fatigue as a continuous measure compared to classification of activity types. Narayana noted that time and frequency analysis typically gave more accurate results, however frequency analysis is not practical with non-periodic data such as the random transition from running to climbing obstacles. Cust[164] showed the volume of research with conventional machine learning and how deep learning is becoming more popular. With the accuracy levels greater than 90% there is an opportunity to reduce the complexity and time intensive nature of data processing by reducing the number of sensors and adopting deep learning to remove the feature extraction phase in the process. This will decrease any inconvenience to the participants and hence enable longer duration activities to be researched in more remote settings. The automatic nature of deep learning over traditional approaches promises a shift to sensor location selection based on participant convenience as features of interest do not need to be known a-priori. Beeck[165] showed that models generally translate across participants and several locations give similar results. Lonini[166] concluded machine learning models should be trained on personal datasets to give the best accuracy, however they also translated well to others. This is an important conclusion as it enables an experimental approach to be used on a single participant to determine if an analytical approach is valid. The collection of ground truth data for labelling is typically performed using: direct observation, video, self-report or by activity prescription during data collection, all of which is not possible in remote scenarios.

The literature review concludes fatigue is a multidisciplinary area of research where a reductionist approach no longer adds to the fields understanding. Research into fatigue needs to be performed in the field in a manner that captures the complexity of multiple interactions taking place with the environment and human physiology, neurology and psychology. Furthermore the literature review identified that little research was done: (i) in an outdoor unstructured environment, (ii) using a single IMU, (iii) generating ground truth data without direct observation or (iv) following voluntary activities and pace. There is no research where fatigue was introduced as a modulating factor into human activity recognition and conversely no research into fatigue where various activities were undertaken in a free living environment. Furthermore rough terrain with obstacles and different surfaces have not been included in previous HAR or fatigue analysis. The use of CNN models with IMU and ECG have been using datasets that do not include environmental noise. This research will investigate how to establish a protocol for data collection that enables data processing, generation of ground truth labelling and supervised machine learning. The research will investigate how to design existing machine learning approaches in the presence of environmental noise and compare the accuracy to previously curated datasets in a laboratory environment. If this research is successful it will result in a protocol and data pipeline that can classify both continuous and one off activities and predict continuous fatigue levels in mountainous terrain over multiple days during voluntary activity with self-pacing and no self-assessment or manual labelling. This will greatly enhance future studies of fatigue and performance in remote field environments.

LINK BETWEEN CHAPTERS 2 AND 3

The literature review identified that fatigue research needed to move from the laboratory to the field where physical, cognitive and psychological effects all interact with the environment and goal. interactions that take place. Furthermore analysis of large complex datasets from wearable sensors should be possible using deep learning tools if the appropriate labels can be generated.

To move fatigue research to the field it was important to define and evaluate a protocol. The protocol required compliance in the presence of distractions and had to supply labelled datasets for later machine learning. The labels needed to be previously validated assessments from a laboratory in order to show the field results were of equivalent accuracy to laboratory results in the literature.

Chapter 3 provides the first field study for the thesis, which focuses on determining which assessments show sensitivity in the field and the compliance for wearables in the field.

The hypothesis for this experiment was that fatigue would be due to motion sickness, a constantly moving yacht and sleep deprivation. Assessments were chosen, a protocol was planned and ethics was approved. The ocean passage was on a 50 foot yacht travelling for 12 days from the east coast of the United States to the Antigua in the Caribbean. This is the normal time of year to deliver a yacht and is at the back end of the northern hemisphere winter so typically gets several days of storms that coincide with crossing the gulf stream which travels in the opposite direction to the northerly artic winds, resulting in large uncomfortable waves. Sailing also includes monotony which makes task completion challenging.

This paper describes a field protocol to measure fatigue and enable machine learning from the dataset. A significant contribution of this paper is the compliance formula that can be useful in developing research protocols to reach a higher protocol compliance by fatigued participants.

3. CALCULATING COMPLIANCE FOR WEARABLES AND COGNITIVE ASSESSMENTS IN THE FIELD: A CASE STUDY DURING A 4-PERSON MULTI-DAY OFFSHORE SAILING VOYAGE

This paper was submitted for publication as:
Russell, B., Hume, P. A., McDaid, A., & Toscano, B. (2020 in review,).
Calculating compliance for wearables and cognitive assessments in the field: A Case Study during a 4-person multi-day offshore sailing voyage. *Journal of Sports Psychology in Action*

Objectives: To investigate the compliance of wearable sensors and tablet based cognitive assessment tests for physical and cognitive fatigue during a 12-day offshore sailing voyage. To derive a compliance formula that enables calculations and trade-offs for wearable devices and protocols under field conditions.

Design: Prospective cohort study.

Methods: Four experienced offshore sailors were assessed using Stroop, Finger Tap Test (FTT), Karolinska Sleepiness Scale (KSS), Borg Scale Rating of Perceived Exertion (RPE) and Pittsburgh Sleep Quality Index (PSQI) on a daily basis for 12 days while wearing a Medtronic BioHarness. Compliance was calculated using a linear deviation from prescribed time-of-day, CTOD, and a binary classifier of compliance for prescribed worn-time, CWT.

Results: Spearman's rank correlation (Rs) for CTOD was variable between participants (Rs -0.58 to 0.79). CWT increased by a factor of 3.3 from the first to second half of the voyage (0.21 to 0.69) based on a linear calculation of compliance with an acceptance threshold of 1 hour.

Conclusions: A formula to calculate for binary and linear compliance to an experimental protocol was assessed. This formula can be useful in future field research when designing assessment protocols.

Keywords: compliance, multi-day, task assessment, wearable sensors, endurance, cognitive, neuromuscular, fatigue, hassle factor, sensor.

Practical Implications:

- A formula was put forward to understand the factors that contribute to compliance in the real-world scenarios, including goal, enjoyment, workload, available resources and number of tasks.
- Results were shown for measured and estimated compliance over subjective assessments, objective assessments and wearables.
- This work allows a systematic approach to design devices and research protocols to improve compliance in real world situations.

Introduction

The aim of the study was to evaluate the compliance of four participants in the use of wearable sensors and tablet-based assessment tests against physical and cognitive fatigue during a 12-day offshore sailing voyage. The voyage was expected to fatigue the participants. It was expected that compliance would be greater for assessments that were less obtrusive (wearing the sensor), or that took less time to complete (questionnaires) compared with more time consuming and possibly technically difficult assessments (Stroop and FTT). The protocol of tests can be considered an additional cognitive load. Data were used to assess compliance to the prescribed time of day to start a battery for self-assessment questionnaires and tests using computer software. A second compliance measure was the cumulative time a wearable sensor was worn each day.

Determining fatigue in multi-day events is important to reduce the likelihood of injury and optimise performance[110,179]. Cognitive fatigue can reduce situational awareness and reduce task adaption based on the real-world changes such as the weather and terrain[2,128,180]. Physical fatigue, if not managed correctly, can lead to injury or not achieving a goal. The combination of unmanaged cognitive and physical fatigue can put a participant in danger [4,10,82]. Subjects in small crew sizes and extreme environments can suffer various cognitive and social deficits[181]

Current measures of cognitive and physical fatigue use active participation by the subject by either answering questions, such as in the NASA TLX test[39,182] or participation in a performance tasks such as Stroop[131],[113]. These tests were chosen as they are validated for cognitive and physical fatigue and hence can allow the measurement of compliance against these criteria as modulation factors; however, they can be problematic for compliance issues when the participant is already occupied undertaking strenuous or dangerous activities in the field. Compliance can be influenced by factors such as workload, fatigue, pain, motivation, culture, personal relationships in the group.

Cognitive tests traditionally are performed in a quiet lab space or simulated environment using equipment such as a computer – therefore they are not easily used in the field. Wearable sensors can be used to measure biomechanics and vital signs; however, they have other labour, comfort and cognitive cost factors in the field.

Noncompliance with assessment tests reduce the accuracy of fatigue modelling and analysis. Ideally a sensor would be used in the field so that active participation in assessment tests was not required. A sensor requires less direct interaction than active testing but does have other factors which affect compliance such as battery loading, donning/doffing and sensor accuracy under all activities. Anecdotally using either active tests or sensors in the field for fatigue studies is that compliance to the test protocol reduces with fatigue.

Methods

The researcher's university ethics committee (AUTEC 17/353) approved all procedures in the study and all sailors gave written informed consent prior to participating in the study.

Four experienced offshore sailors (14, 51, 52, 54 years, 76 ±19 kg, 1.75 ±0.15 m) undertook an offshore sailing trip from the east coast of the United States (Annapolis) to the Caribbean (Antigua) over a distance of 2,600km for 12 days. Participant occupations included three business executives and a student. Work shifts for each participant were agreed and fixed for the same 4-hour period during each 24-hour period to determine sleep periods[70,72]. A daily diary was kept with results from the self-assessments and cognitive tests, sleep and work shift times. Cognitive and physical loads were expected as part of the activity.

The test battery included self-assessment of physical condition via questionnaires and cognitive tests via computer. The test battery was prescribed to be completed once per day for each sailor just after waking to account for circadian rhythms. Participant A voluntarily completed tests more than once daily to determine inter-day variation and determine training effects.

Self-assessment of physical condition via questionnaire ratings were collected by using an XL spreadsheet (Microsoft, Redmond, USA) for; quality of sleep using the Pittsburgh Sleep Quality Scale[154,183] („sub section 1"subjective sleep quality", 0=very good to 3=very bad), sleepiness using the Karolinska Sleepiness Scale[115] (1=extremely alert to 0=very sleepy, great effort to stay awake), mood alertness state using a Visual Analog Scale, VAS[184] (Yes/No answers to each question; stomach awareness, stomach discomfort, headache, dizziness, vomiting), physical exertion using the Rating of Perceived Exertion scale[24] (6=no exertion to 17=very hard) and motion sickness symptoms using the Graybiel Diagnostic scale[185] (rating of none, mild, moderate, severe for each question; nausea, salivation, pallor, sweating, drowsiness, temperature). Cognitive and neuromuscular tests via computer included the Stroop test[130,131,186] and Finger Tap Test[71,112] operated on an iPhone. Data was recorded to the participants diary after each test[113]. Participants were asked to wear a BioHarness[187] vital sign monitoring device around their chest for 21-hours per day. This period was chosen to allow battery recharging and data download. The data logged on the device included electrocardiogram (ECG) waveforms sampled at 250 Hz and accelerometer data in three axes at 100 Hz (x-axis=vertical, z-axis= anteroposterior and y-axis=mediolateral). Heart Rate, calculated by the device, with a value above 35 beats per minute was used to determine if the sensor was worn by a participant. It is expected that participant B and C had low compliance scores for wearing the device due to ill fitted sensors not detecting a valid cardiac signal. These data were downloaded to a laptop daily. Summary data (including heart rate and activity) were calculated by the device and output at one second intervals. The participants were given an information document on the protocol and instructions on the various tests and donning/doffing the sensor. Several practice sessions took place the day before the protocol commenced.

Compliance (C) to the assessment tests and vital sign monitoring were calculated as compliance to test protocol start time-of-day (C_{TOD}) and compliance to sensor wear time (C_{WT}) versus a required threshold of one hour per day. Compliances per day per participant were calculated using a binary compliance equation (1) when a pass-fail criteria was required and calculated using a linear compliance equation (2) with thresholds equal to 1-hour and 2 hours.

Binary compliance, for a single task where, T_A is the actual task value, and T_{TH} is the threshold value.

$$\boldsymbol{C = \begin{Bmatrix} 1, if\ T_A > T_{TH} \\ 0, if\ T_A < T_{TH} \end{Bmatrix}} \qquad \textbf{(3-1)}$$

Linear compliance for a single task were, T_A is the actual task value, and T_{TH} is the threshold value.

$$\boldsymbol{C = \begin{Bmatrix} 1\ , if\ T_A > T_{TH} \\ \dfrac{|T_{TH} - T_A|}{T_{TH}}, if\ 0 < T_A < T_{TH} \end{Bmatrix}} \qquad \textbf{(3-2)}$$

C_{TOD} was the time absolute difference between the prescribed and actual start time of the test protocol for each day measured in minutes. For each participant a single score was derived from each questionnaire and cognitive test across the 12 test days resulting in 84 data points per participant.

C_{WT} used total periods of 10 minutes throughout the day that the device was worn, determined from valid heart rate values over 35 bpm inside the 10-minute time blocks.

Spearman ranked correlation, Rs, were performed for each test across the 12 days for each participant, with p values calculated using a t statistic (degree of freedom = 10. A t statistic was used with a degree of freedom equal to 10. A significance level of 0.05 was used to report correlations between data series. Python with the SciPy library was used for these calculations. Daily Mean Absolute Error across the cohort was calculated for compliance to a start time and total minutes worn for the wearable sensor. Inter series correlations (such as time of day error vs FTT) used actual error to preserve direction. For example if the test was started one hour early the error would be reported as -1 hour.

Cohort linear fit was performed using cohort MAE value for each day. A linear regression with p value was calculated using Excel's functions (Pearson,() RSQ(), T.DIST.2T() and t= (r*sqrt(n-2))/(sqrt(1-r^2)).

Results

Compliance to test protocol start time-of-day (C_{TOD}) was calculated Figure 3-1 shows (a) start time error values with the ±1-hour threshold shown and (b) the resulting compliance per day calculated against the threshold of one hour.

Figure 3-1a shows the test protocol start time-of-day error time in hours (prescribed to actual times to start the test protocol*)* over the voyage for all the sailors. The ±1-hour either side is shown as a horizontal line. Most participants were compliant within one hour of the prescribed time for the first six days of the trip (<2 hours deviation) and then deviated in the second six days (>8 hours on average). The mean was calculated using mean absolute error (MAE). Figure 3-1b shows the calculated compliance from the error times using equation (3-2).

The start time error linear regression (R^2 0.49,p=0.01) indicated a moderate group correlation with a linear trend.

Table 3-1 shows the linear and binary daily calculations for each participant's compliance over each day. Results were calculated per participant for the entire multi-day test. The cohort mean was 33.3%. When the threshold was increased from 1-hour to 2-hours the participants mean compliance increased to 65% (Participants percentages at 2-hours were A 50%, B 50%, C 75% and D 65%).

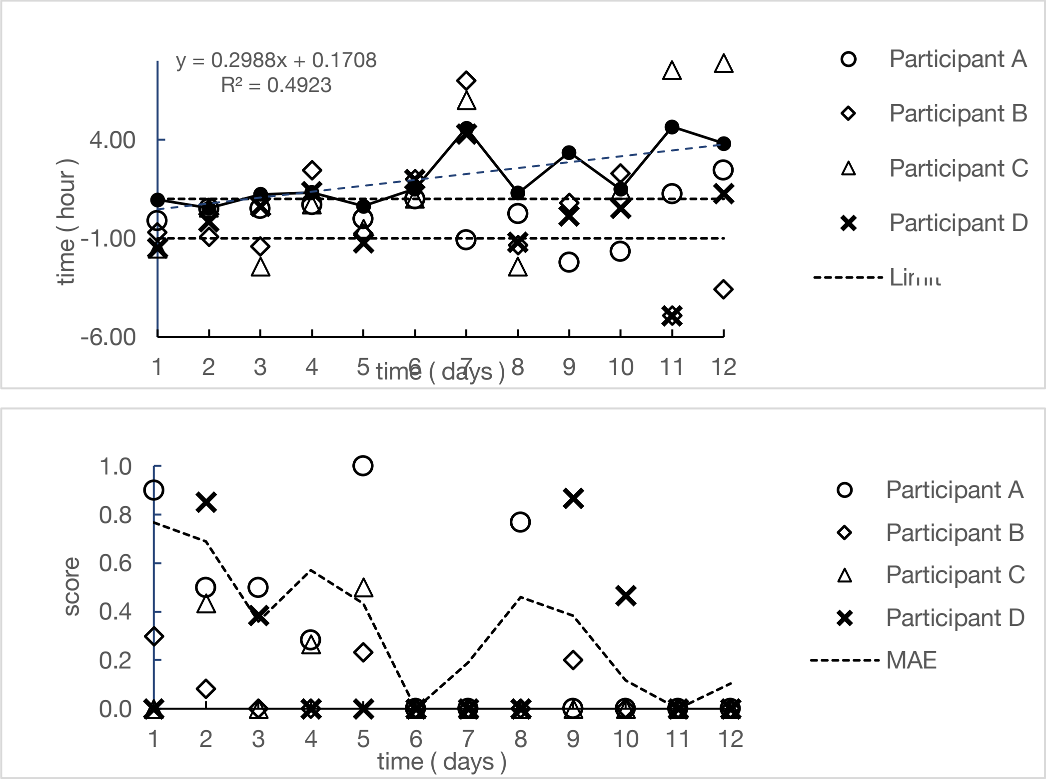


Figure 3-1 (a) Time error plot for test time during each day, (b) Compliance for Time of Day score calculated per day.

Table 3-1 Compliance results for time-of-day start time for questionnaires per day per participant using the binary compliance equation and the linear compliance equation with the threshold equal to 1 hour.

	Time of day (decimal time, hours)				Linear Compliance (1-hour threshold)				Binary Compliance (1-hour threshold)			
	Participant				Participant				Participant			
Day	A	B	C	D	A	B	C	D	A	B	C	D
1	9.90	11.30	11.48	10.50	0.9	0.3	0.0	0.0	1	1	0	0
2	10.50	11.08	13.57	11.85	0.5	0.1	0.4	0.9	1	1	1	1
3	10.50	10.60	10.58	12.62	0.5	0.0	0.0	0.4	1	0	0	1
4	10.72	14.47	13.73	13.38	0.2	0.0	0.3	0.0	1	0	1	0
5	10.00	11.23	12.50	10.77	1.0	0.2	0.5	0.0	1	1	1	0
6	11.00	14.00	14.00	14.00	0.0	0.0	0.0	0.0	0	0	0	0
7	8.92	19.00	19.00	16.30	0.0	0.0	0.0	0.0	0	0	0	0
8	10.23	10.67	10.58	10.82	0.8	0.0	0.0	0.0	1	0	0	0
9	7.78	12.80	2.73	12.13	0.0	0.2	0.0	0.9	0	1	0	1
10	8.35	14.30	14.45	12.53	0.0	0.0	0.0	0.5	0	0	0	1
11	11.27	7.08	20.53	7.08	0.0	0.0	0.0	0.0	0	0	0	0
12	12.45	8.42	20.90	13.28	0.0	0.0	0.0	0.0	0	0	0	0
Total									6	4	3	4
Percentage compliance (%)					33	7	20	21	50	33	25	33

Compliance to sensor wear time (C_{WT}) was calculated. Figure 3-2a shows the daily sensor worn time in minutes over the voyage for all the sailors. Most participants were not compliant within the 1-hour threshold for the first six days of the trip and then deviated significantly in the second six days. Figure 2b shows the calculated compliance sensor time within the 1-hour threshold.

The compliance of wearing the sensor for 1-hour per day was very low (37%). The last half of the protocol showed better compliance which may be due to familiarity with wearing the sensor. Worn time calculation is based on valid heart rate data, as the crew were asked to wear the sensor for one hour before each test time. The particularly low numbers may be due to loss of signal, the sensor not being turned on or the data download failing.

Figure 3-2c shows the results of the compliance calculations for both time of day and sensor worn as a daily mean across the cohort. There was insufficient evidence to show a significant correlation between participants for compliance to either start time or wear time.

The relationship between compliance and fatigue was assessed. C_{WT} increased by a factor 3.3 from the first half to second half of the research period (0.21 to 0.69) based on a linear calculation of compliance with an acceptance threshold of 1 hour. Stroop reduced by 38% over the period for the cohort (R^2 0.79, p 0.0001). Stroop, and protocol day was strongly correlated for significance level of 0.05 (Rs -0.66 to -0.85, p 0.00 to 0.02). FTT and protocol had a strong correlation for subject C (Rs -0.62, p 0.03) showing it was monotonic. Sleep Quality and KSS were strongly correlated (Rs 0.61, p 0.04) for subject A who was jet lagged and on night shift.

All four participants showed reduced Stroop results (Figure 3-3) indicating decreased cognitive function (Rs -0.66 to -.85, p = 0.00 to 0.02). Participant A voluntarily carried out the test more than once a day, and the test closest to protocol time-of-day was used for Stroop results. Subject A was well practiced at the tests, so a training effect was not expected for that individual's data.

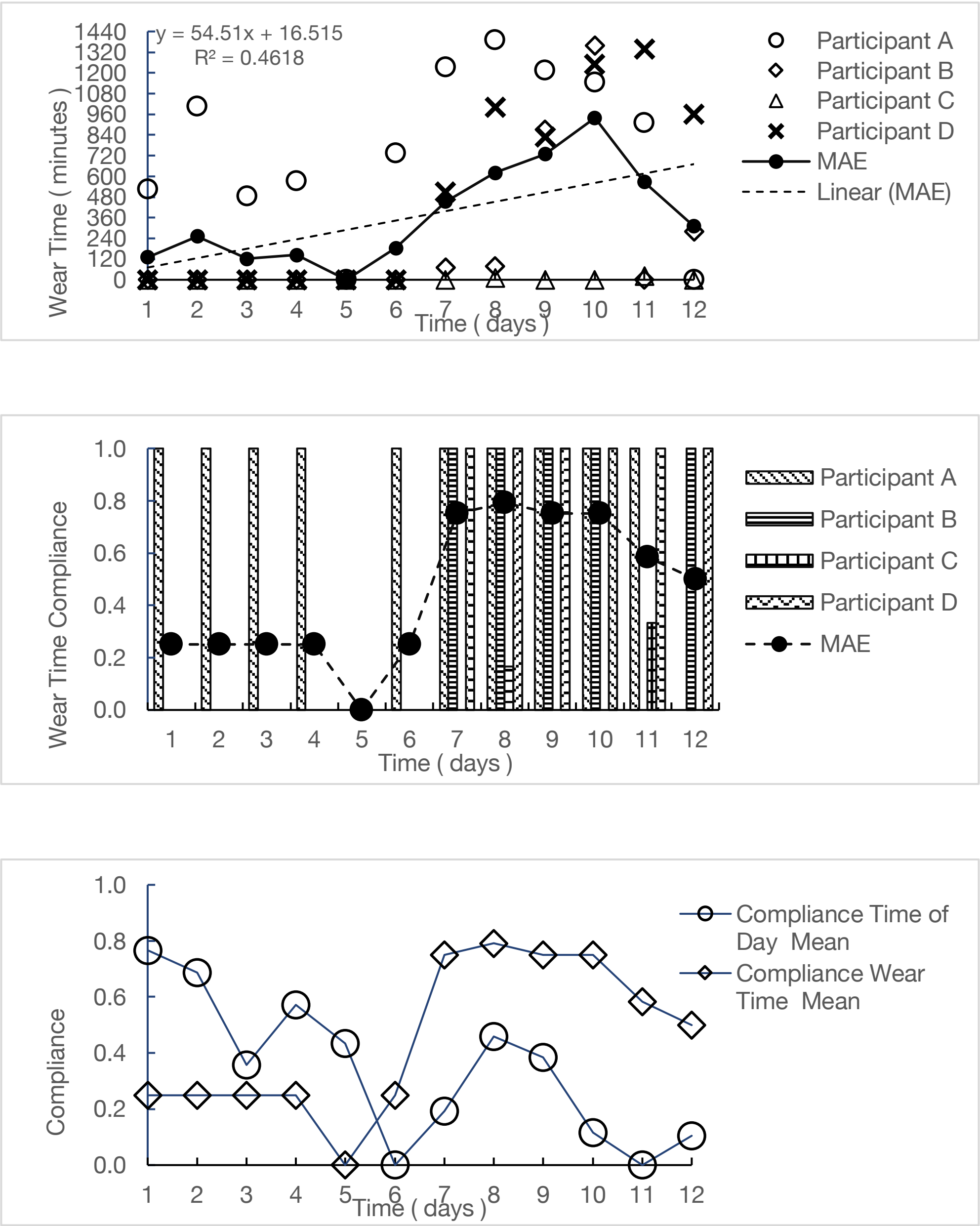
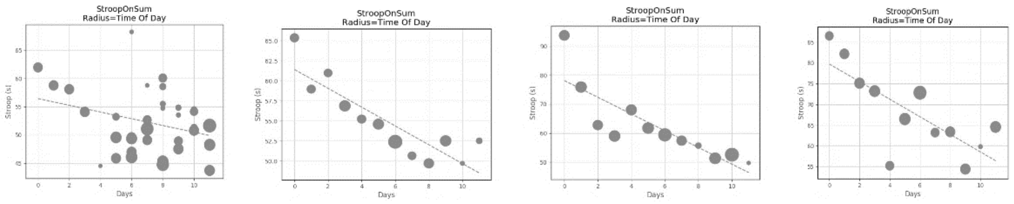
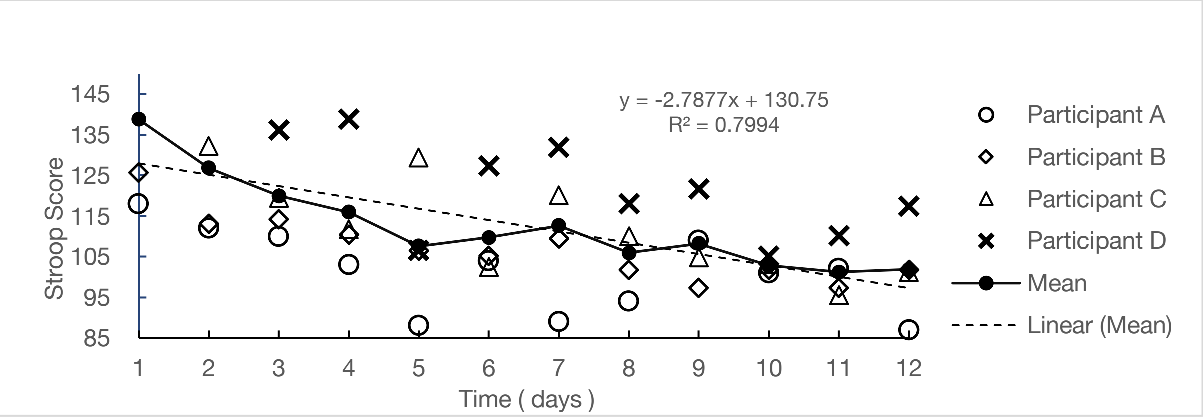


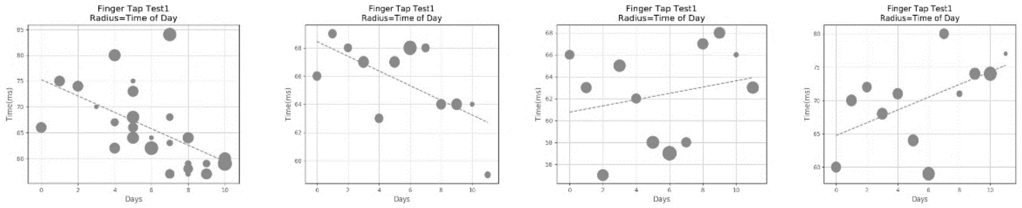
Figure 3-2 (a) Total minutes the sensor was worn per day by the four participants (b) Compliance for sensor wear time per day, 1 hour = 100% compliance (c) Compliance from two methods.

The Finger Tap Test (FTT) results (Figure 3-3)) with respect to time of day showed the daily cohort mean time changed by less than 2% over the test period, indicating that neuromuscular function was constant for the cohort. Individual results varied over the protocol by 25 seconds (mean=66, SD=5.85, min=55 s, max=80 s). FTT and KSS were strongly correlated for three participants A,B,D Rs -0.6 to -0.7, p= 0.01 to 0.04). This makes sense that sleep quality is related to sleepiness.



Participant A Participant B Participant C Participant D





Participant A Participant B Participant C Participant D

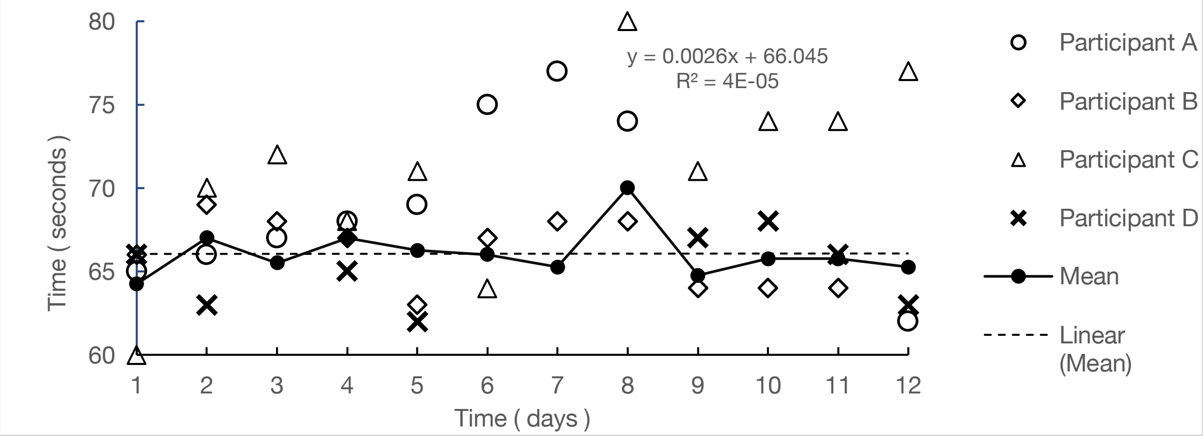


Figure 3-3 (a) Stroop scores for individual participants with circular radius representing the time of day the test was undertaken, (b) Stroop scores over the research period combined with cohort mean, (c) Finger Tap Test results for individual participants shown time of day with FTT result as the circular radius, (d) FTT results over the research period combined with cohort mean.

The Karolinska Sleepiness Scale (KSS) was generally larger (more fatigued) in the early hours of the morning and did not show any trends across days. KSS changed by 4% of the total range (mean 3.9, SD 2.0) giving an insignificant daily change compared with variations within a 24 hour period. Participant A was jet lagged and had the work shift from 12 am to 4 am. Participant D had a busy high-pressure job and he stated the sailing trip was relaxing compared to his normal life.

The Borg scale Rating of Perceived Exertion (RPE) had a daily cohort range of two RPE units (mean 9.7, SD 1.4) with a minimum when participant B reported very low numbers. RPE and wear time compliance was strongly correlated for three participants A,B,D (Rs 0.6 to 0.77, p 0.003 to 0.04).

The Pittsburgh Sleep Quality Index (PSQI) increased (indicating worse sleep) on day 5 during the worst weather for the rip. PSQI was strongly correlated with Stroop (Rs -0.61, p 0.03) for participant A probably due to a combination of jet lag and night shifts from 12am to 4am.

Discussion

To enable accurate measurement of cognitive and physical performance in the field over multiple days it was first necessary to determine compliance to self-assessment questionnaires, cognitive tests and the use of wearable devices to monitor vital signs. It may be possible to model this compliance in order to predict compliance and gain insights into how various input factors can affect the outcome. This may lead to a useful work flow that can assist in product design for wearable sensors.

The sensor was worn more at the end of the protocol (days 7-12), possibly due to increased familiarity, and the questionnaire testing dropped off, possibly due to the cognitive load reducing motivation. This led to a lower initial compliance in Table 3-1. The questionnaire start time compliance dropped off towards the end of the test period possibly due to reduced motivation of the cognitive load. The last day showed a high compliance to C_{TOD} which showed all four participants were diligent in preparation for arriving at land.

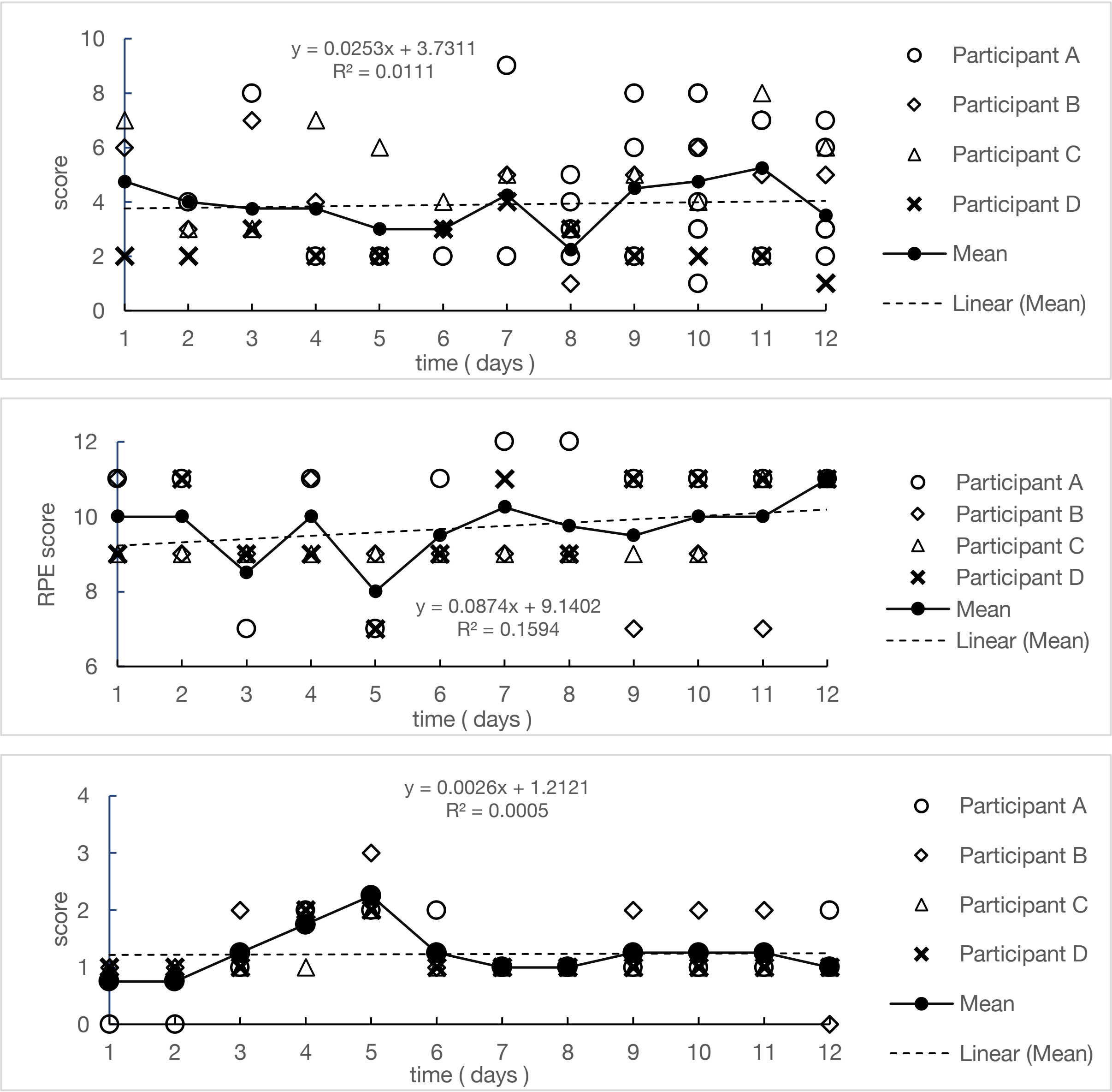


Figure 3-4 Summary self-reporting indices across the cohort (a) Karolinska Sleepiness Scale, (b) Rating of Perceived Exertion, (c) Pittsburgh Sleep Quality Index

It was expected that compliance would be greater for assessments that were less obtrusive (wearing the sensor), or that took less time to complete (physical condition questionnaires) compared with more time consuming and possibly technically difficult assessments (Stroop and FTT). It was also expected that environmental effects such as weather and sea condition changes, and night versus day work shifts, could have an effect on the sailor’s fatigue.

The sailing voyage did not provide the expected fatigue due to unseasonably good weather resulting in only slightly decreasing fatigue scores from both Stroop and FTT. The compliance across the group dropped between the first and second half of the 12 day period both with test time compliance and increased for sensor worn compliance. The lack of sensor wear compliance for the first six days was unexpected and was likely due to the definition based on valid heart rate data rather than actually wearing the sensor. It is instructional that compliance based on sensor data may in influenced by invalid physiological data being used as a worn detection.

The participant variability differed on a daily basis but trended in the same direction from compliance to non-compliance from day-1 to day-12 with the sensor compliance increasing and more onerous diary questions and cognitive tests becoming less compliant.

Compliance to the start time protocol and sensor worn protocol was low. Each individual's response to sleep cycles and motion sickness was different. Time zone acclimatization may have affected one participant (10 hours' time zone offset arriving the day before the protocol started). Further study is required in a scenario where all participants' activities are synchronized. The assessment protocols used a different iPhone application for each test and required the same demographic data to be added each day resulting in repetitive tasks. Further research is needed with the possible inclusion of the following recommendations: 1) using a streamlined single application with less testing per instance (shorter test durations), 2) more tests per day to allow higher resolution data throughout the day and 3) increase compliance with a sensor that can be worn 100% of the time without charging or down loading and should connect to a tablet or iPhone app to confirm valid vital signs are being logged. Video of the assessment test protocol activities for post study compliance assessment is needed to confirm the protocol is adhered to.

Based on the experiences in the sailing voyage data, we now propose that a Compliance Index can be calculated. Compliance can comprise multiple tasks, for example, if a device requires charging, $C = 1$ if the device was successfully charged to full. Example sub tasks may include the plug was inserted correctly, the wall plug switch was not turned on and the laptop screen was left open (to enable USB ports to stay active). Total compliance can be summed based on the compliance of N tasks;

$$C_T(N) = \frac{\sum_{i=1}^{N} C_i}{N} \qquad \textbf{(3-3)}$$

Where

C_i is compliance of a single task, between 0 and 1,

N is total number of tasks

We measured compliance and calculated estimated compliance based on the success of individual tasks. The estimated Compliance ($\tilde{C}$) can be derived from task specific costs and available resources. Resources in this context can refer to available calories, time or cognitive energy based on how fatigued a person is. Cost can refer to spent calories, time or perceived effort. $\tilde{C}$ may be time varying and subject specific. $\tilde{C}$ decreases exponentially with the total number of tasks and is modulated by the Hassle Factor (H), which increases cost per task and decreases with available resources.

Model based estimate of Compliance, $\tilde{C}$

$$\boldsymbol{\tilde{C} = \frac{1}{exp\left(\frac{N_s - 1}{H}\right)}} \qquad \textbf{(3-4)}$$

Where:

Ns = number of sub tasks to complete a goal

H = Hassle Factor Index

The Hassle Factor Index (HFI) is a measure if a task contributes to a desired goal and the ratio of available resources versus the cost to complete that task. Units of resources and cost can include: perceived cognitive or physical load, time or energy, volume or weight. Hassle factors may include number of sensors, time to complete the tasks, number of sub steps, distraction from other tasks or goals. A wearable can add a hassle factor based on location, weight, size, interference during activities, skin abrasion, impact risk, data download steps, battery charging, snag risk, donning, doffing, perception of others when using the device. Component contributors are modulated by two factors; fun factor (F_n), and goal factor (G_n). F_n is 1 if the task is fun and 0 if the task is boring. G_n is 1 if the task contributes to the mission goal and 0 if task obstructs the mission goal.

Note the term $K/(F_n.G_n)$ can be termed as motivation to perform the task. This modulation factor can be viewed as a factor influenced by the central governor theory [30]. Intuitively it represents how enjoyable and how related to a goal an activity is. The sum or costs over resources can be viewed as work load. Hence Hassle Factor Index can be viewed as task specific motivation multiplied by work load.

The Hassle Factor Index for a given task is given by:

$$HFI(n) = \frac{K}{F_n G_n} \sum_{i=1}^{N} \propto_i \frac{C_i}{R_i} \qquad \textbf{(3-5)}$$

Where:
$\propto_i$ = weighting factor for each ratio.
C_i = costs of performing the task
R_i = resources available for the task
K = constant used to scale probability of compliance between 0 and 1, typically = 10
F = 'fun factor', positively motivates a participant to complete the task. 1=very enjoyable, 0=boring
G = 'goal factor', representing achieving a goal. 1 = highly desirable, 0 = undesirable

For the example of the sailing data, this is what we propose

$$HFI = \frac{1}{F_n G_n} .(\propto_S \frac{C_S}{R_S} + \propto_W \frac{C_W}{R_W} + \propto_Z \frac{C_z}{R_z} + \propto_B \frac{C_B}{R_B} + \propto_D \frac{C_D}{R_D} + \propto_E \frac{C_E}{R_E})$$

(3-6)

Where

C_s = social influences such as wearing obvious versus hidden from view.
R_s is social influences such as how accepting others are to the activity or wearable.
C_w = weight of unit or device
R_w = total weight carrying capacity of the subject
C_z = size of sensor or device for the activity
R_z = total volume carrying capacity of subject

C_b = battery charge time
R_b = battery charging time available
C_d = data download time
R_d = data download time available
C_e = cognitive effort of the task
R_e = cognitive attention available for the task

An example for Hassle Factor Index results for participant A in the sailing voyage was calculated as 27.5 with 25 total number of steps. The Estimated Compliance was calculated as 0.42. Further evaluation is needed of these proposed variables for compliance to tests in the field during multi-day events.

Various factors influence the total Hassle Factor Index for various tests. BioHarness is dominated by data download time followed by time to don and doff the sensor. Stroop is dominated by cognitive load which is logical given it is a cognitive load assessment. Finger Tap Test is weakly affected by cognitive load. Self-assessment tests are dominated by cognitive load and social effects.

Table 3-2 Compliance results for wearable and assessment tests measured versus estimate calculations

	Participant				Group	
	A	B	C	D	Mean	SD
Wearable Compliance						
Compliance - worn 1 hour per day	0.8	0.4	0.0	0.5	0.43	0.33
Probability of Compliance						
BioHarness worn for 1 hour	0.4	0.4	0.4	0.4	0.4	0
Questionnaire Compliance						
Daily Start Time	0.5	0.3	0.25	0.33	0.35	0.11
Probability of Compliance						
Stroop	0.2	0.2	0.2	0.2	0.2	0
FTT	0.4	0.4	0.4	0.4	0.4	0
Questionnaire (PSQI, KSS)	0.7	0.7	0.7	0.7	0.7	0

Table 3-2 shows the various compliance values that can be calculated by participant. The wearable compliance for participants is shown against the calculated probability of compliance based on the Hassle Factor formula in equation (6). The wearable had a probability of compliance (P(wearable)=0.4) which is an equivalent probability to FTT. Stroop is seen a low value (P(Stroop) = 0.2) compared to FTT (P (FTT)=0.4) and the various questionnaires (P(KSS) = P(PSQI) = 0.7). When calculating the HFI and probability of compliance all tasks were rated as low for “contribution to goal”. Wearing the sensor and answering questions were rated fun factor as low or “boring” whereas the FTT and Stroop were rated high or “fun to perform”.

Table 3-3 Calculation tables for Hassle Factor Index for various types of assessment with explanation

BIOHARNESS	Goal	Fun		Units	Notes
K			0.1		Scaling factor to determine reasonable C
Gn			0.1		Wearing the device did contribute to goal.
Fn			0.2		Neither fun or boring.
$=K/G_n.F_n$			5		
	Cost	Resource			
	C_i	R_i	C_i/R_i		
Time	60	600	0.1	minutes	Time required to wear device each day.

Cognitive load	0.1	0.8	0.125		No concentration required.
Social	0.2	0.9	0.22222		Could not notice the device being worn.
Weight	0.05	10	0.005	kg	Very lightweight
Size	1	5000	0.0002	cm^3	Size was small.
Battery charging	60	600	0.1	minutes	Charging took 1 hour.
Data download time	120	600	0.2	minutes	Inconsistent results took multiple attempts
Encumbrance	0.1	1	0.1		Didn't notice wearing the device
HFI			4.26211		
Number of sub tasks			5		Don and doff the device, check the LED is on and flashing, charge and download
STROOP	Goal	Fun		Units	Notes
K			0.1		Scaling factor to determine reasonable C
Gn			0.1		Did not contribute to goal
Fn			0.9		Fun to play
$=K/G_n.F_n$			1.11111		
	Cost	Resource			
	C_i	R_i	C_i/R_i		
Time	2	600	0.00333	minutes	2 minutes out of a 10-hour period
Cognitive load	1	0.2	5		Requires concentration
Social	0.8	0.9	0.88889		Not embarrassing
Weight	0	1	0	kg	Tablet was not carried so not relevant
Size	1	5000	0.0002	cm^3	Tablet was not carried so not relevant
Battery charging	30	600	0.05	minutes	Tablet required charging
Data download time	1	600	0.00167	minutes	Tablet recorded automatically low hassle
Encumbrance	0.1	1	0.1		Tablet on table in galley so no problem
HFI			6.71565		
Number of sub tasks			12		Enter name and age, perform 10 times
FINGER TAP TEST	Goal	Fun		Units	Notes
K			0.1		Scaling factor to determine reasonable C
Gn			0.1		Testing did not contribute to goal
Fn			0.9		Fun to play
$=K/G_n.F_n$			1.11111		
	Cost	Resource			
	C_i	R_i	C_i/R_i		
Time	2	600	0.00333	minutes	2 minutes out of a 10-hour period
Cognitive load	0.5	0.6	0.83333		Requires concentration

Social	0.2	0.9	0.22222		Not embarrassing
Weight	0	10	0	kg	Tablet was not carried so not relevant
Size	1	5000	0.0002	cm^3	Tablet was not carried so not relevant
Battery charging	30	600	0.05	minutes	Tablet required charging
Data download time	1	600	0.00167	minutes	Tablet recorded automatically low hassle
Encumbrance	0.1	1	0.1		Tablet on table in galley so no problem
HFI			1.34528		
Number of sub tasks			2		Take part for left and right hand
SELF ASSESSMENT	Goal	Fun		Units	Notes
K			0.1		Scaling factor to determine reasonable C
Gn			0.1		Testing did not contribute to goal
Fn			0.3		Boring
$=K/G_n.F_n$			3.33333		
	Cost	Resource			
	C_i	R_i	C_i/R_i		
Time	1	600	0.0083	minutes	2 minutes out of a 10-hour period
Cognitive Load	0.5	0.2	4		Requires concentration
Social	0.2	0.9	0.2222		Not embarrassing
Weight	0.05	10	0.005	kg	Tablet was not carried so not relevant
Size	1	5000	0.0002	cm^3	Tablet was not carried so not relevant
Battery charging	60	600	0.1	minutes	Tablet required charging
Data download time	60	600	0.1	minutes	Tablet recorded automatically so low hassle
Encumbrance	0.1	1	0.1		Tablet on table in galley so no problem
HFI			15.119		
Number of sub tasks			6		Answer questions, caffeine, sleep time, KSS, Sleep quality, Mood, PMVT

Limitations

The small number of participants in this study limits the statistical power, a larger crew would induce different environmental loads with respect to; work load, social stressors, perceived risk, sleep, nutrition, cognitive and physical loads. The various back grounds of the crew also adds statistical noise but is representative of the diverse back grounds and work environments that an offshore crew will experience.

Conclusion

The objective of this study was to investigate the compliance to a protocol that included a wearable chest sensor and various cognitive assessments. A formula to quantify compliance was proposed and assessed. Our findings suggest that binary and linear compliance can be calculated to enable insight into the design of protocols.

LINK BETWEEN CHAPTERS 3 AND 4

The ocean passage field trial described in chapter 3 resulted in significant insights into protocol design for: (i) compliance, (ii) researcher work load in field conditions, (iii) sufficiently frequent labelling to detect short term changes in the subject fatigue, (iv) assessment types to allow post data collection choice of assessments as labels, (v) data analytics with small numbers.

The conclusion from this study determined that a different protocol was required in the transition from laboratory to field. A protocol was required that had (i) a predictable load, (ii) broad assessment coverage to allow inter inter-subject variation, (iii) frequent assessments to capture short term fatigue variation, i.e. hourly not daily (iv) assessment tools to reduce repetitive tasks as this introduced errors and lead to irritation from monotony.

During the ocean passage the crew members shared an iPad to perform several assessments using different applications. The motivation to enter repetitive data, such as name and date of birth, reduced rapidly. A conclusion from this field study was that each subject should have their own iPad with only one app incorporating all assessments. This allows personal data to be entered once and assessments to be carried out efficiently.

This lead to the development of one app that was allocated to a single person to alleviate entering demographic and had all the appropriate Likert questionnaires and assessments in one user experience.

In chapter 4 the questions are:

(i) Can activity labels be determined with direct observation from location and acceleration?
(ii) Can human activity recognition (HAR) be determined with a machine learning model from acceleration and ECG when trained against manual labelled data, complex terrain, obstacles and a fatiguing subject?
(iii) Is the HAR model comparable in accuracy to less noisy laboratory datasets?

Any machine learning model is limited in accuracy by the accuracy of the dataset labels. The software development to semi automate the labelling process was an unexpectedly large task taking several months requiring multiple unreported data science experiments. The volume of data adds delays in processing time between iterations and manually checking results includes observing sections of 5 second data snippets in 11.8 million data points. An example of an activity transient label including slope, surface type and activity is where a subject runs over flat gravel up to an obstacle, walks two steps over flat dirt, climbs over the obstacle, walks several steps over long grass on flat ground and continues running up a clay track.

4. MOVING THE LABORATORY INTO THE MOUNTAINS: A PILOT STUDY OF HUMAN ACTIVITY RECOGNITION IN UNSTRUCTURED ENVIRONMENTS

This paper was submitted for publication as:
Russell, B., Hume, P. A., McDaid, A., & Toscano, B. (2020 under review,). Moving the lab into the mountains: A pilot study of human activity recognition in unstructured environments. *IEEE Transactions on Biomedical Engineering (TBME)*.

Background: Automated human activity recognition (HAR) has reached a high level of accuracy in the lab using sensors and machine learning. To date most data gathered for HAR has been orchestrated in a structured environment with low levels of fatigue to enable a protocol that controls data gathering under specific and repetitive activities. If HAR is to transition into the field a methodology is required to gather high quality in-field data which incorporates random variation from the environment, participant fatigue and develop models that can attain high accuracy with this environmental noise.

Aim: To create and validate a field-based data collection and assessment method for human activity recognition in the mountains.

Methods: The protocol generated a labelled dataset of various long term field-based activities including run, walk, stand, lay and obstacle climb. Activity was voluntary so transitions could not be determined a priori. Terrain variation included slope, crossing rivers, obstacles and surfaces including road, gravel, clay, mud, long grass and rough track. Fatigue levels were modulated from rested to failure. The dataset was used to train a deep learning convolutional neural network (CNN) and the HAR results were compared to a lab-based dataset.

Results: The experimental results showed that the model performed well with tuned hyper parameters. The trail run dataset had 3,829,759 samples with 5 features included repetitive activities and single instance activities which required hyper parameter tuning to reach an overall accuracy 0.978 with a minimum class precision for the one-off activity 0.802.

Conclusions: To the authors knowledge this study demonstrated the first successful HAR in a mountain environment. A robust and repeatable protocol was developed to generate a validated trail running dataset when there are no observers present and activity types changed on a voluntary basis. The experimental results show the machine learning model performed well compared to lab equivalents (accuracy 97.8% vs 97.7% for trail vs lab).

Introduction

HAR has been well researched in the lab using both ambient and wearable sensors[188],[189] including repetitive and single activities [190, 191]. Chen[192] achieved 93.8% accuracy using a smart phone for eight low intensity activities in a controlled environment where the users self-labelled activity with an app. Narayanan et al recently reviewed 53 studies and found with either 1 or 2 sensors accuracy ranged from 62% to 99.8 % for posture and activity type in controlled environments[163]. It has been shown that gait is different when an athlete trains on stable versus unstable surfaces[193] and variations in fatigue[74]. Ambient sensors include cameras and load cells which require the subject to perform in a fixed location. Abnormal Human Activity Recognition, abHAR, has been researched using video[190]. Inertial measurement units (IMU) are wearable sensors that can measure acceleration, rate of rotation and magnetic fields. IMUs have the advantage of being wireless, battery powered and small. Field based studies are more suited to IMU sensors where the subject may be out of sight. The number of sensors worn by a subject is a trade-off between the researcher requiring more detail and the subject complying while carrying out a task. In multiday scenarios battery power and memory capacity start to constrain the number of sensors that can be used. Mountain trail running is long, fatiguing and requires light weight equipment leading to a requirement for a single IMU sensor. Machine learning techniques have shown good accuracy when analysing IMU sensors once the data is labelled allowing supervised learning approaches.

Related Work

Various activities have been studied using IMUs including; activity of daily living[194] with the UCI50 and WISDM datasets[161], factory workers[195], food preparation[196], tennis, snowboarding, weight lifting, rugby and running[197,198]. Johnson et al[199] used cameras and load cells with deep learning to predict ground reaction forces. Xu et al[200] model complex activities with 52 channels from 7 IMUs. Buckley et al[171] modelled a binary classifier for fatigue using 3 IMU sensors to classify fatigue or non-fatigue. The protocol measured athletes on two 400 metre runs on a track separated by a Beep test until a self-reported Borg[24] scale exceeded 18. Buckley shows that IMU data can be used to determine biomechanical states over variation in fatigue. However, the labelling of data was trivial given the protocol determined when the two binary states took place during data collection.

Accelerometer Data From The Field Over Various Surfaces

Various running surfaces with changes in slope, texture and obstacles are expected to give an increased to gait variability. Figure 4-1 illustrates how the accelerometer waveforms change for running up and down a hill (slope between 8 and 10 degrees) and how this changes when travelling over a hard road surface vs a soft uneven track surface. Zero crossing statistics are summarized in Table 1-1 to highlight the variation in timing across different conditions showing the standard deviation increases with the transition from hard smooth to software textured terrain. The question this paper investigates is whether a deep learning model can perform at sufficient accuracy given the data perturbation that occur in trail running compared to lab based data.

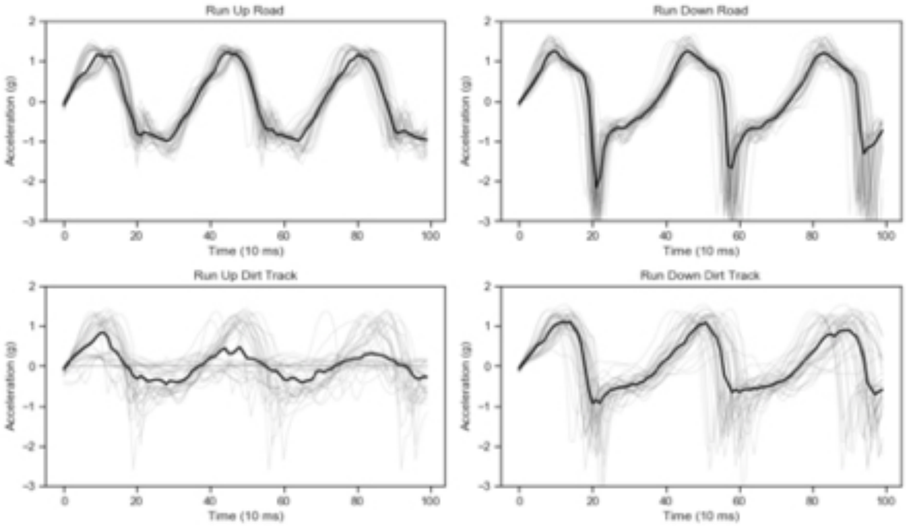


Figure 4-1 Acceleration waveforms for running up and down slopes over various terrain (road, track)

Table 4-1 Vertical axis zero crossing times for two activities and three terrain surfaces

	Zero Crossing Times	
Condition	Mean (s)	SD (s)
Run Up		
Road	0.365	0.013
Track	0.381	0.030
Run Down		
Road	0.352	0.014
Track	0.475	0.272

Labelling Data

Validation of activity, surface characteristics, and distance travelled in the lab occurs naturally with the observation of the researcher to the prescribed activity and the surface of lab floor likely not being noted due to its homogeneity. Alternatively users can label their own data when a smart phone app is used for labelling and accelerometery[192]. Activities in a remote field environment are self-selected, surface types change with location and speed is unpredictable being determined by surface and fatigue levels. The user is occupied and cannot label data. This study outlines a protocol to address the challenges of non-observed field-based research to enable lab equivalent data validation and HAR detection accuracies.

In this study, we propose a framework and work flow to calibrate a trail running track in the mountains in order to gather unsupervised data and assign verifiable labels in order to train a deep learning model. The data tuning and model optimization will be described to train a CNN and compare results to training with a previously published lab-based WISDM dataset[203].

Machine Learning

Deep learning automatically learns features in the data, replacing manually determined features and potentially using a larger number of features[195]. Many studies have investigated deep learning models using previously published datasets using multiple sensors in a controlled environment. Wang[188] recently surveyed approaches in HAR, noting deep learning techniques are replacing traditional pattern recognition techniques for repetitive activities such as running or walking, but unable to determine activities such as having coffee. Nweke[191] reviewed deep learning algorithms for human activity using wearable sensors and found they out performed traditional machine learning models using manual features. He also concluded that variation in the data is required to make the model generalizable.

Various machine learning models have reported good results for HAR with deep learning models showing good results with their ability to automate the extraction of features. Wang surveyed the various models used for HAR including; convolutional neural networks (CNN), recurrent neural networks (RNN) including the subset of long term short term memory (LSTM)[201], deep belief networks (DBN) , restricted Boltzmann machines (RBN), stacked autoencoder (SAE) and hybrid combinations of the above[188].

CNN has several advantages over other models including: local dependency, scale invariance and being capable of processing on multiple processors. Local dependency can take advantage of HAR signals being related when close in time. Scale invariance means patterns can be determined as they change over time, cadence or amplitude. RNN has also been used for HAR and its derivative Long Short Term Memory[201] with the limitation of not being able to use parallel processing and having stability problems.
Window sizes in the range of 2 to 10 seconds have been reported as giving best results for HAR[194,202]. Wang[194] et al discussed the trade-off between speed and recognition performance. Banos[202] et al showed that window sizes of less than 2 seconds gave good detection performance. There are varying requirements for single instance activities (e.g. climbing a fence) compared to repetitive activities (e.g. running, walking). Overlap can allow an increased classification accuracy of times series data. A strategy must be chosen to determine the label to be used for entire window when each sample is individually labelled. We will experiment with various window sizes and window labelling methods to optimize for this dataset.

The number of samples used in this work will be greater than the WISDM dataset[203] by a factor of three. Datasets in the lab often have multiple subjects to address generalizability and statistical significance. This work will present a novel dataset, experimental protocol and HAR model to determine the validity of the approach replacing intersubjective variation with uneven terrain (slope and surfaces), obstacles and fatigue. Firstly a mountain trail is calibrated with waypoints and labels for any transition in terrain texture (concrete, gravel, grass, mud, rocks, rivers) and slope. The participant activity is self-selected by the subject in the mountains, unobserved and the subject is increasingly fatigued (both physical and mental) throughout the protocol. Single instance activities (climb gate) are included with repetitive activities (lay, sit, walk, run)

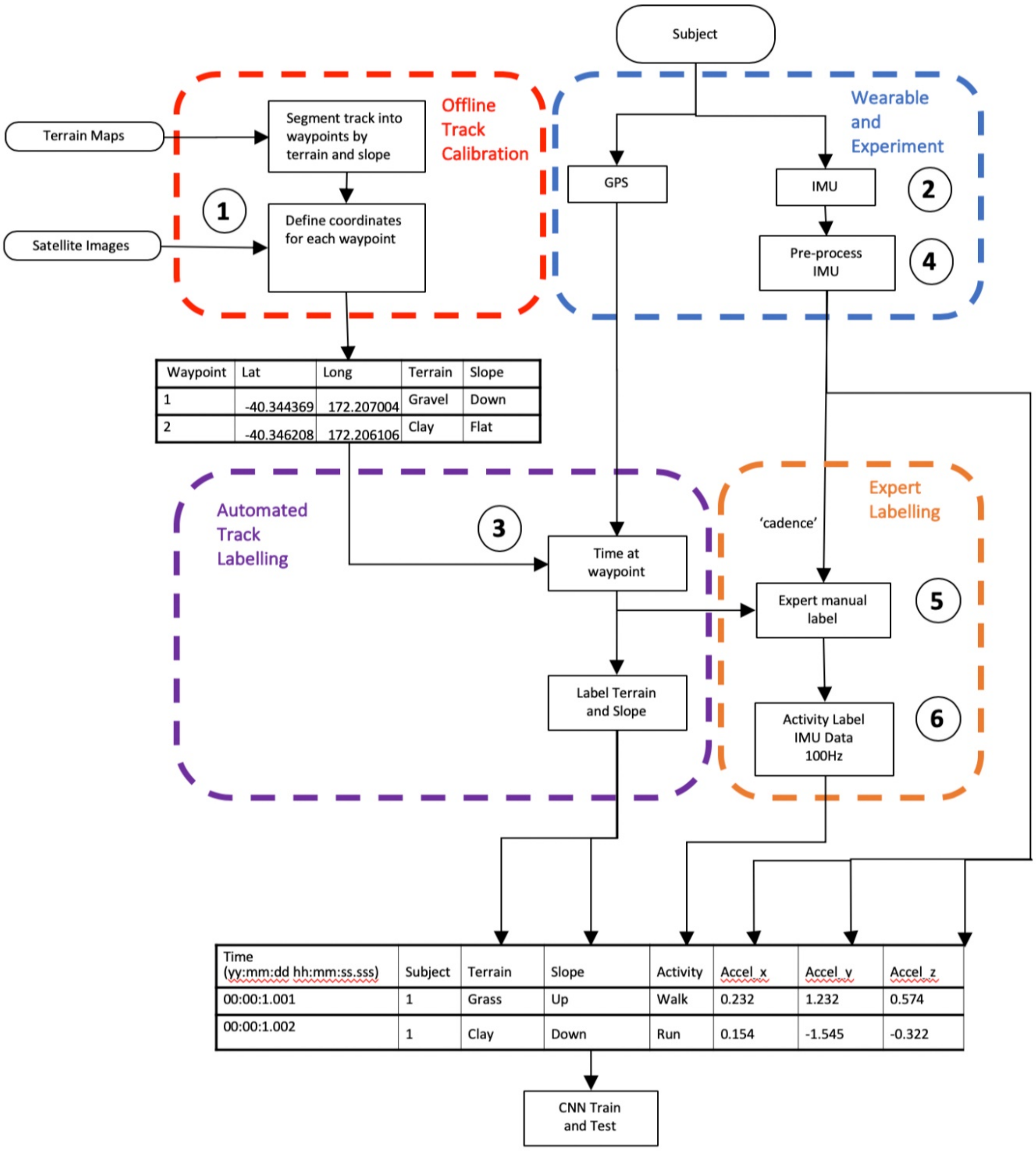
The primary contributions of this work are firstly the capability of calibrating an outside track for field based experiments and secondly the development of a HAR model for mountain track.

Methods

Ethics

The researcher's university ethics committee (AUTEC 18/412) approved all procedures in the study and the participant gave written informed consent prior to participating in the study.

An overview of the methods is described in Figure 4-2 with pseudo code describing the steps in Table 4-2



Waypoint	Lat	Long	Terrain	Slope
1	-40.344369	172.207004	Gravel	Down
2	-40.346208	172.206106	Clay	Flat

Time (yy:mm:dd hh:mm:ss.sss)	Subject	Terrain	Slope	Activity	Accel_x	Accel_y	Accel_z
00:00:1.001	1	Grass	Up	Walk	0.232	1.232	0.574
00:00:1.002	1	Clay	Down	Run	0.154	-1.545	-0.322

Figure 4-2 Data pipeline from sensor data to training deep learning model

Table 4-2 Pseudo code of trail calibration, protocol, data analysis and deep learning model

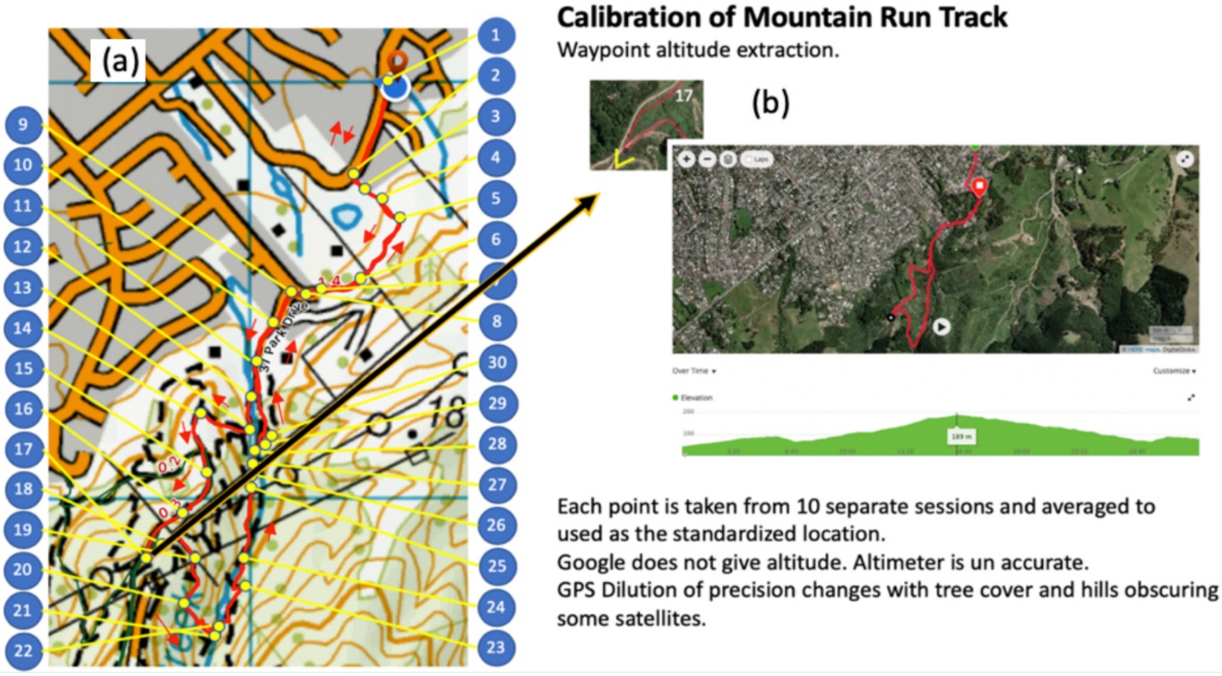
	Process
1.	Define a course by waypoints, *WP*(n, x, y, z) where x, y, z are (latitude, longitude and altitude) and n =1,2..N. Each waypoint defines a change in Terrain (concrete, grass, clay,..), Slope (down steep, down, flat, up, up steep) and Object (fence, gate, river, stairs). Define the same start and end waypoint, *WP*(1)=*WP*(N) .
2	Define the location along the course, *GPS*(t, x, y, z), where t_s = 1,2,.. seconds and x, y, z are latitude, longitude and altitude. Define the 3 axis IMU data along the course, *IMU*(x, y, z, t_{ms}), , where t_{ms} = 1,2,.. milli seconds GPS(t, x, y, z) is the location over time of the participant. Align time stamps between sensors by calculating time offset. Manual correction is performed by an expert observing the acceleration waveforms.
3.	Collect GPS and IMU data from participant during experiment.
4.	Define the time at each waypoint *T*(n) = min (*WP*(n, x, y, z) - *GPS*(t, x, y, z)).
5.	Normalise data using maintaining inter axis scale IMUmin = min (IMU(IMUx, IMUy, IMUz)), IMUmax = max (IMUx, IMUy, IMUz) IMUnormal(x, y, z) = (IMU(x, y, z) – IMUmin) / (IMUmax-IMUmin)
6.	Pre-process including calculating *CADENCE*(t) = zero_crossing *IMU*(y, t_{ms})
7.	Observe *IMU*(x, y, z, t_{ms}) and *CADENCE*(t) to label activity at each time step. Allocate ACTIVITY(t) based on observation.
8.	Allocate labels terrain, slope where IMU(terrain, slope, x, y, z, t_{ms})
9.	Allocate activity labels ACTIVITY(t)
10.	Train the CNN model using IMU(x, y, z, t) against ACTIVITY(t)

Trail and Calibration

A 3.8km mountain trail with a total of 194 m vertical elevation and duration of 25 to 35 minutes for a fit healthy adult was selected. The route was chosen such that it had various terrain and slopes and with the same start and end location.

The trail was divided into segments defined by waypoints, where each segment had a feature change, such as terrain or slope Figure 4-3(a).

The course was segmented with a GPS location and altitude for the start of each segment. A segment was started if a feature changed including; slope (up, flat, down), surface texture (concrete, gravel, mud, grass) or obstacles (river, gate, stairs). Small obstacles such as an open gate, rocks, branches or holes in the track did not require a new segment. Video (GoPro Hero 4, Garmin Kansas, USA) was used to validate terrain type. Each waypoint was located on Google Earth satellite view and topographic map data (source: www.linz.co.nz, www.garmin.com) using vegetation or ground features for registration and manually recorded to six decimal places equating to 10 cm accuracy Figure 4-3(b). A comma separated variable file was generated with columns for waypoint number, waypoint name, latitude, longitude, altitude, terrain slope in degrees, terrain texture.



Calibration of Mountain Run Track

Waypoint altitude extraction.

Each point is taken from 10 separate sessions and averaged to used as the standardized location.
Google does not give altitude. Altimeter is un accurate.
GPS Dilution of precision changes with tree cover and hills obscuring some satellites.

Figure 4-3 (a) Track calibration with segmentation waypoints and (b) ground features recorded for reference. (source: www.linz.co.nz, connect.Garmin.com)

The subject wore a BioHarness[204,205] around the chest (Medtronic, Minnesota, USA) for acceleration data (sample rate 100 Hz, x-axis=vertical, z-axis= anteroposterior and y-axis=mediolateral) and Garmin Forerunner GPS watch on the wrist (sample rate 1 Hz, horizontal accuracy 6 m) to assist with labelling.

The protocol was broken into hourly segments which can be viewed as three sections, firstly physical load followed by cognitive load and finally a small rest period before restarting at the top of the hour. The physical load performed is trail running with obstacles (25 to 35 minutes depending on level of fatigue) immediately followed by a 15-minute cognitive fatigue test battery (
Figure 4-4).

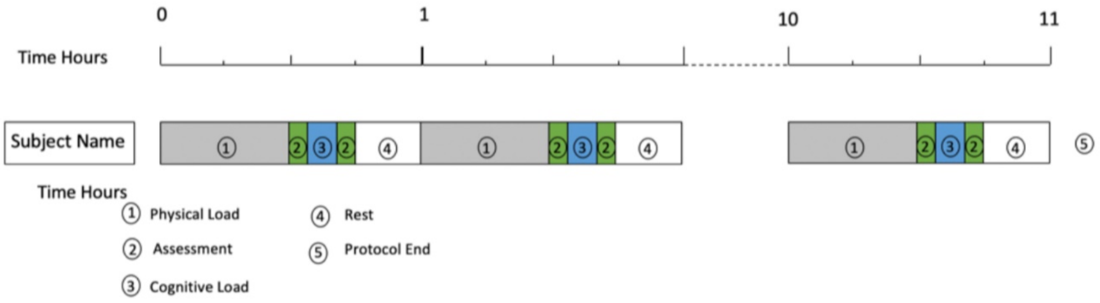


Figure 4-4 Protocol description

A custom app (COGNI) which incorporated the individual test batteries of Stroop[131,109], Finger Tap Test[112], Trail Making A[144,206], Trail Making B, Spatial Memory, was performed on an iPad Pro (Apple, Cupertino) and a custom implementation of the Multi Attribute Test Battery (MATB)[207] was performed on an Apple MacBook Pro (Apple Cupertino , USA) running in Python. The COGNI testing was completed twice with the MATB assessment in between Figure 4-4

For determining ground speed the participant was instructed to “go as fast as possible AND repeat trail run 24 times over a 24 hour period. This protocol was part of a larger study into fatigue and the results of the various tests are not reported in this paper. The participant could not continue due to fatigue after 11 circuits. This demonstrates the extreme levels of fatigue that are included in the dataset, i.e. from no fatigue through to complete fatigue.

Dataset Preparation for HAR

The aim of the data processing was to generate a dataset with columns for; subject, timestamp, acceleration vertical, acceleration sagittal, acceleration lateral and activity label. Where subject is the person’s anonymized number, timestamp is date and time to an accuracy of microseconds, acceleration channels are direct from BioHarness and the label is the activity label with a resolution of 1 second.

Location of the subject on the track was determined by analysing the GPS to derive the time of closest proximity to a waypoint. The time at each waypoint was determined when the subject to waypoint proximity inflected from approaching to leaving.

Activity labelling started with feature extraction of cadence in steps per minute based on a zero crossing of the vertical acceleration (100 < walk < 150, run > 150). Activity type was further refined based on knowledge of the protocol sequence, the track, waypoint proximity times and observation of the acceleration waveforms, where Figure 4-1 shows an example of running up to a gate with a transition to walking before climbing the gate, walking then running.

Acceleration data was normalized. Where the same minimum and maximum for all three axes are applied to maintain scale relationships. Total time running was 6 hours 40 minutes.

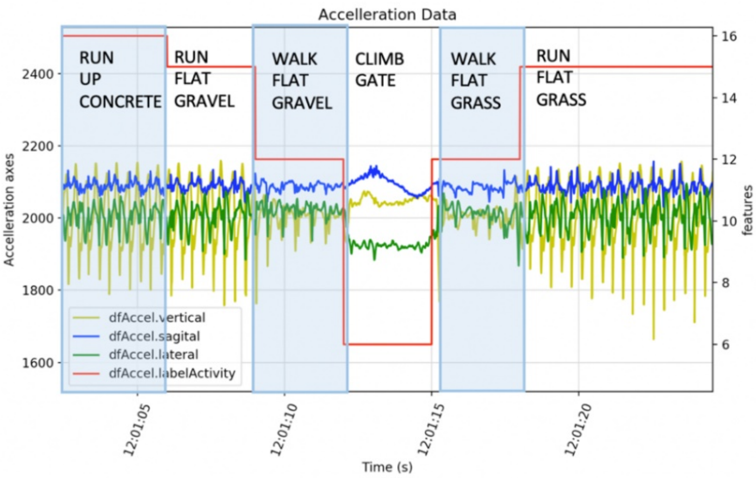


Figure 4-5 Time series example of one-off activity "climb gate" versus repetitive data "run" and "walk"

Features refer to variations such as slope or surface texture. Labels refer to the final activities used to train the model. Feature columns were combined to derive multimodal features in a single column, such as terrain derived slope and activity to enable ‘run-up’ vs ‘walk down’. The 100 Hz acceleration data was allocated an activity label per row by time synchronizing segments of samples between each waypoint using the asynchronous labelled location data.

Data was reduced from 3,829,759 samples to 3,341,184 samples by removing data that did not have the following labels (lay, sit, climb gate, walk, run) and during rest periods where activity was not prescribed or recorded. A comparison of the dataset labels with the WISDM dataset is given in Table 4-3.

Table 4-3 Dataset sample count by label compared to a public dataset

Label	Trail Running	Label	WISDM
Sit	1,065,100	Sit	59,939
Walk	741,599	Walk	424,398
Run	1,438,302	Jogging	342,176
Climb Gate	26,099	-	
Lay	70,100	-	
-		Stand	48,395
-		Down Stairs	100,427
-		Up Stairs	122,869
Total	3,341,184	Total	1,050,048

CNN models require segments of data to be input during training. Time series data with n samples were divided into segments of W samples wide with S overlap leading to D rows of data, equation (4-1).

$$\boldsymbol{D} = \frac{(n - W)}{W - S} \qquad \textbf{(4-1)}$$

Each data point in the window can have different labels however a decision is made for one label per window using majority voting with a threshold, TH_{MV}, treated as a hyper parameter in the training model. Majority voting can assist when activities change and a dataset includes two activities whereas the CNN classifier is required to choose one class. Labelled windows were excluded if the majority label vote did not exceed a threshold. For example $MV_{TH} = 0.4$ majority voting would exclude a window if 40% of the samples were not identical.

Time Series data were transformed into an array (D,W,F) with D rows, W samples wide with number of features, F, equal to 3 from the accelerometer axis (x, y, z). Randomized train test split of 0.33 was selected using sklearn in python.

Tuning of the window size and window labelling rejection ratio was performed to optimize accuracy between repetitive activities and one-off activities.

Deep Learning Model

For classification, a CNN topology was used, as shown in Figure 4-6. This network consists of 3 separate 1D convolutional networks for each acceleration axis joined by a dense layer to achieve a multivariate classification. Each axis includes two convolutional 1d layers with filter size 64, kernel size 3 and ReLu activation. This is followed by a drop out layer set at 50% for generalizability and a max pool layer with pool size 0.5. The 3 separate channels are combined in a flatten layer followed by a dense layer using a ReLu activation and finally a softmax activation function to give a probability density function for each class. Learning used an epoch of 100 with batch size of 50. The classification accuracy was calculated for each label as a multivariate analysis.

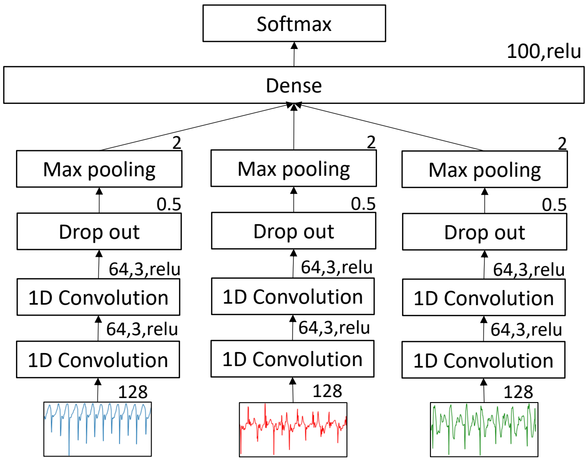


Figure 4-6 Structure of CNN

Results

Experiment One

The majority voting threshold, MV_{TH}, for each window was tested from a value of 0.20 to 0.95 for labels (lay, sit climb gate, walk, run), epoch 100, batch 50, stride 256, overlap 128, windows 26201. Minimum accuracy was at MV_{TH} 0.20, accuracy 0.973, rejected windows 0. The Accuracy increased monotonically and was flat above MV_{TH} 0.8, accuracy 0.982, rejected windows 494. As such we chose 0.2 and 0.8 in the following experiment.

Experiment Two

A second experiment was run with MV_{TH} 0.2 and 0.8 with a range of window sizes, W and overlap, S, shown in Table 4-4. The highest overall accuracy 0.982 for W 256, S 128 and MV_{TH} 0.8, however the ‘Climb Gate’ precision was highest for W 128, S 64 and MV_{TH} 0.80 with overall accuracy dropping by 0.04 to 0.978. The trade-offs can be seen in Figure 4-7.

Table 4-4 Results for Trail Run for CNN over stride and label rejection ratio, F=3, Labels=5

Window Size W	Overlap S	Ratio MV_{TH}	Window Accept	Window Reject	Accuracy	Precision				
						Lay	Sit	Climb Gate	Walk	Run
16	8	0.2	417648	0	0.972	0.996	0.970	0.674	0.964	0.979
16	8	0.8	417044	604	0.973	0.996	0.972	0.708	0.964	0.980
32	16	0.2	208823	0	0.973	0.991	0.968	0.702	0.972	0.981
32	16	0.8	208313	510	0.977	0.989	0.976	0.737	0.973	0.981
64	32	0.2	104411	0	0.973	0.972	0.975	0.664	0.967	0.980
64	32	0.8	103979	432	0.974	0.980	0.974	0.580	0.974	0.981
128	64	0.2	52205	0	0.973	0.988	0.978	0.702	0.954	0.982
128	64	0.8	51721	484	0.978	0.972	0.975	0.802	0.966	0.988
256	128	0.2	26102	0	0.973	0.929	0.984	0.700	0.956	0.981
256	128	0.8	25608	494	0.982	0.977	0.977	0.771	0.980	0.989
512	256	0.2	13050	0	0.970	0.921	0.990	0.773	0.939	0.976
512	256	0.8	12613	437	0.981	0.987	0.973	0.625	0.977	0.991
1024	512	0.2	6524	0	0.953	0.971	0.964	0.455	0.884	0.986
1024	512	0.8	6135	389	0.980	1.000	0.976	0.500	0.988	0.980

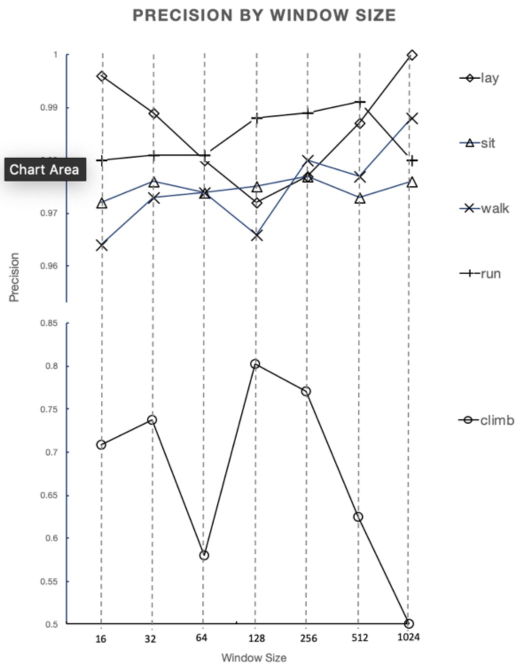


Figure 4-7 Results for Trail Run for CNN over stride and label rejection ratio, F=3, Labels=5

Discussion

A protocol was presented to calibrate an outside trail running track with novel data segmentation and labelling methods. To our knowledge this is the first time a dataset has been collected in an unstructured mountain environment and validated with a deep learning model. This work will allow further research in applications outside of the lab in rough terrain and with voluntary movement over time.

A CNN deep learning model was used as a multi variate HAR classifier on a dataset from a trail run. The protocol included voluntary transitions between activities, various obstacles, variation in surface texture, terrain slope and subject fatigue. The 3 axis 100 Hz acceleration data was labelled by activity with a temporal resolution of one second. The resulting time series was sliced into windows where majority voting assigned a single label per window. The resulting array was passed to a CNN multivariate classifier and trained to learn activity from the acceleration signals. Increasing the value of the majority voting threshold improved accuracy for all 5 classes of activity with no improvement for threshold values over 0.8. The dataset included repetitive activities and single instance activities which required hyper parameter tuning to reach an overall accuracy 0.978 with a minimum class precision for the one-off activity 0.802.

The trail run dataset has 3,829,759 samples with 5 features and 1 participant. The comparison WISDM dataset by Kwapisz[208,203] et al 2010 has 1,098,207 samples over 6 attributes, with 33 participants. The trail run variations are due to terrain and fatigue whereas the WISDM dataset variation is derived by multiple subjects. Accuracy with the same CNN structure resulted in an accuracy of 0.977. It should be noted that the WISDM dataset min precision was 0.928 whereas the one off activity “climb gate” achieved an accuracy of 0.80 and hence the trail running CNN model accuracy shows that deep learning HAR can be performed accurately in the field with appropriate protocol to enable labelling. There is a trade off in window size for repetitive activities and one off activities where it is recommended to use a window size of 128 to ‘climb gate’ over 80% precision as the over accuracy only reduces by 0.4%.

From these results it can be confirmed terrain variation and obstacles can be incorporated into a protocol outside of the lab with self-selected activity when using a single sensor and a deep learning model.

This work was limited to one participant and further work is recommended to extend the models to multiple participants using the protocol described. Sleep deprivation can be induced by reducing the hourly physical load which in turn allows a longer time period for the protocol. GPS accuracy in tree covered deep valleys was reduced and having video on the person for every run would speed up manual labelling, computer vision could be used to recognize waypoints.

The effectiveness of the protocol is encouraging as it enables translational research to be under taken in environments closely related to where they will have the most impact. For example soldiers work load and training on multiday missions can determine activity types to adjust team work load and maximise speed and reduce the chance of injury. Adventure sport people can gain insight into pace and activity types to improve training and reduce the chance of injury. This protocol enables performance analysis for the likes of selection in military applications or activity recognition for remote workers in dangerous environments.

Conclusion

This paper presents a framework which includes a track calibration protocol, data collection protocol and data analysis pipeline. The framework will allow a trail running track in the mountains to be calibrated to facilitate an activity protocol with labelled data when the activities types were self-selected over time. This dataset has similar classes of activity to lab data but includes intentionally induced modulation in gait using the environment and fatigue as the stimulus.

A multi variate CNN deep learning model was implemented on time series data with hyper parameter tuning to maximize overall accuracy 0.978 and individual class precision (0.801 – 0.988).

In the experiments conducted during the study, we confirmed that a deep learning model can accurately classify activities from an accelerometery dataset from a trail run with modulated; terrain, slope and subject fatigue. We also evaluated the same model on a previously published WISDM dataset[203] where the data were gathered in the lab. The results showed similar accuracy and precision results.

In future work, we will conduct further studies over multiple days to include sleep deprivation to the fatigue protocol and increase the number of subjects to assess the generalizability of the model. Further work could also assess if ensemble models enable variable window sizes to attain higher precision for one off activities.

LINK BETWEEN CHAPTERS 4 AND 5

Chapter 4 established that a field environment could be calibrated to enable a protocol for human activity recognition. The experiment also showed that human activity recognition could be modelled and predicted accurately using a CNN classifier unobserved, over rough terrain and with varying degrees of fatigue.

Chapter 4 takes the chapter 5 and changes the CNN model to add ECG as a different data source with different sample rate. Now it uses a regression model outputting a continuous scale. The work looks at long term trend fatigue and also short-term fatigue changes within the protocol.

This chapter shows that cognitive and physical fatigue can be accurately modelled using a single sensor and an AI model with a new dataset from the field. Accuracy results are comparable to previous laboratory results even though the person was self-pacing, fatigued, travelling over an uneven surface compared to a laboratory floor and was deciding from moment to moment on the activity type while not following a prescribed protocol.

This work is exciting in that it enables research in biomechanics, physiology and cognition outside the laboratory plus scale independent of the number of researchers available. The protocol can be reused for other research and shows the potential to eliminate Likert type tests and let the performer get on the task at hand while being analysed.

5. PREDICTING FATIGUE IN MOUNTAIN TERRAIN WITH A SINGLE SENSOR AND MACHINE LEARNING

This paper was submitted for publication as:
Russell, B., Hume, P. A., McDaid, A., & Toscano, B. (2020 under review,). Moving the lab into the mountains: A pilot study of human activity recognition in unstructured environments. *International Journal of Sports Physiology and Performance*

Aim: To determine whether an AI model and single sensor measuring acceleration and ECG could model cognitive and physical fatigue for a self-paced trail run.

Methods: A field-based protocol of continuous fatigue repeated hourly induced physical (~45 minutes) and cognitive (~10 minutes) fatigue on one healthy participant. Physical load was a 3.8km, 200 m vertical gain, trail run with acceleration and electrocardiogram (ECG) data collected using a single sensor. Cognitive load was a Multi Attribute Test Battery (MATB) and separate assessment battery including the Finger Tap Test (FTT), Stroop, Trail Making A and B, Spatial Memory, Paced Visual Serial Addition Test (PVSAT), and a vertical jump. A fatigue prediction model was implemented using a Convolutional Neural Network (CNN).

Results: When the fatigue test battery results were compared for sensitivity to the protocol load, FTT right hand (R^2 0.71) and Jump Height (R^2 0.78) were the most sensitive while the other tests were less sensitive (R^2 values Stroop 0.49, Trail Making A 0.29, Trail Making B 0.05, PVSAT 0.03, spatial memory 0.003). Best prediction results were achieved with a rolling average of 200 predictions (102.4 s), during set activity types, mean absolute error for ‘walk up’ (MAE_{200} 12.5%) and range of absolute error for ‘run down‘ (RAE_{200} 16.7%).

Conclusion: We were able to measure cognitive and physical fatigue using a single wearable sensor during a practical field protocol including contextual factors in conjunction with a neural network model. This research has practical application to fatigue research in the field.

Introduction

Why we need to measure physical and cognitive fatigue in the field

Measures of physical and cognitive fatigue are needed in the field to improve performance and help improve safe participation in outdoor environments.

Physiological and cognitive fatigue in field environments directly effects performance as a person modulates decisions based on contextual input to maintain resources[209]. Operational safety and its relationship to fatigue has been investigated with pilots[128,179] motor vehicle drivers[2,40,210–213] firefighters[214,215] and shift workers[216]. Physical fatigue relates to reduced force, endurance, level of effort, strength, speed and coordination[217]. Levels of performance may be modulated by physical load, sleep, nutrition and psychological factors based on mission duration, pain, levels of perceived exertion[4,30,218,219], intensity and time on task[220]. Hill[16] won the Nobel prize for his work on skeletal muscle and maximum oxygen uptake.

The interaction of central fatigue and motivating factors have been modelled in various forms: Borg's[221] Rating of Perceived Exertion (RPE); Millet's[222] Flush model for pacing strategies in ultra-marathons; Noake's[219] central fatigue model; and Venhorst's[31] bio-psycho-social model.

Cognitive fatigue can be viewed as a combination of goal, adaption and reward trade-offs including the energetic requirements to achieve a goal[33,34]. Performance psychology [36,223] describes performance as recalling ones knowledge, skills and abilities during an event. Cognitive and physical fatigue have a complex interaction of over lapping redundant systems[224].

How we can measure physical and cognitive fatigue in the lab and the field

Mental and physical fatigue have been researched in the lab[58,225]. Fatigue studies have used computer interaction[56], accelerometery, electroencephalogram (EEG), electrooculography (EOG)[57], electromyography (EMG) and electrocardiograph (ECG)[4,58–60], however, these techniques are not always practical in a field setting.

A minimum number of sensors is required to reduce the burden of wearing and maintaining devices to perform fatigue testing in challenging multiday events while not distracting the operator from their mission tasks. A review of sensors used for measuring occupational fatigue[226] showed the most effective sensors were heart rate and accelerometery. A review of physical and cognitive fatigue has shown a relationship of heart rate and accelerometery data caused by changes in muscle activity, proprioception and gait[69,73,74]. Gait has been shown to change physical performance with increased mental fatigue[4,14,213], goals[86] and reduced executive function[35].

Assessment of performance and fatigue have been studied[128] with multiple sensors and neural networks. However, they have not been validated in the field, with noise sources such as terrain, slope and obstacles. Enoka[28] noted that lab based experiments such as maximum voluntary contractions (MVC) result in task dependency that do not translate into field performance. The reduction of separate effects do not equate to overall performance. The only way to determine performance reductions from fatigue is to measure the response to loads in the field.

Traditional machine learning with feature extraction has been used in applications such human activity recognition[194], however this approach assumes the features of interest are known and calculatable. Deep learning uses models which automatically determine feature morphology and significance in the data which may not be observable with traditional statistics and data analysis. Deep learning has been used for areas such as wakefulness detection with accelerometery and ECG[177] and fatigue estimation by Gordienko[178] showed positive results with a repetitive exercises in the gym. Recurrent neural network (RNN) and long short-term memory (LSTM) are often cited as the preferred models for time series data[160]. Convolutional neural networks (CNN) have also been used for time series data[161] and do not suffer from the stability issues of RNNs while enabling parallel processing which is not possible with RNN type models. CNN models have shown good performance on physiological time series data for emotion classification[167], and mental fatigue[168] using EOG which is not generally practical in field operations with high levels of activity. Accelerometery has been shown to be affected by cognitive fatigue[68].

Aim

The aim of this study was to determine whether an AI model using data from a single sensor measuring acceleration and ECG could accurately model cognitive and physical fatigue for a continuous trail run modulated by voluntary temporal activity, terrain and load.

Methods

Ethics

The researcher's university ethics committee (AUTEC 18/412) approved all procedures in the study and the participant gave written informed consent prior to participating in the study.

Protocol – Physical and Cognitive Load and Performance Assessments

A protocol was required that included self-paced running in an unstructured mountain environment and standard performance assessments with no distractions.

The protocol was developed using physical and cognitive loads in excess of a participants' critical power[27] to induce fatigue. A one-hour period of fixed load was repeated until the participant voluntarily ended participation. No restart was allowed. Physical load was provided by a trail run (3.8km, 200 m vertical gain), cognitive load was provided by 10 minutes Multi Attribute Test Battery (MATB)[207] (Figure 5-1).

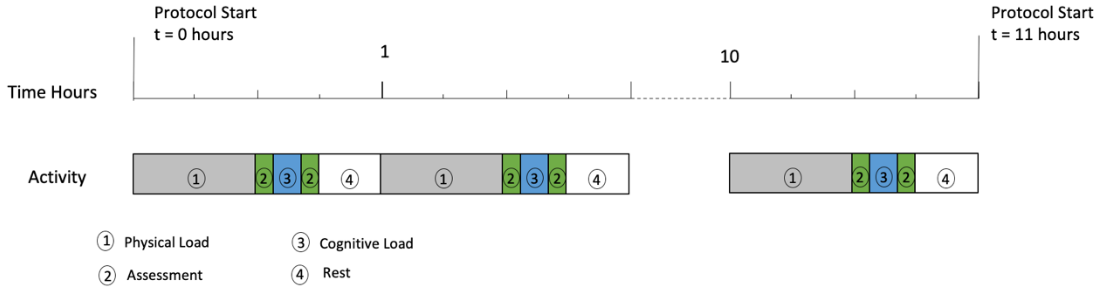


Figure 5-1 Fatigue Protocol

A goal was set as 100km distance, 5200 m (17,000 feet) total climb and 26 hours' time in order to address motivation[30] and psychological perception of pain[227]. The course was prescribed to cover various slope angles and terrain types (concrete, gravel, dirt, grass, boulders) and obstacles (trees, river, gate, fence) and not require active navigation for safety under fatigue and reduced decision making capacity[228]. Speed was rewarded by earlier completion of the hourly protocol resulting in a larger rest period per hour.

Performance assessments where completed on an iPad Pro (Apple Cupertino) using a custom application implementing tests built with Apple Research Kit[229]. The battery of assessments was chosen that have previously shown sensitivity to the protocol loads and fatigue related diseases. These included assessments used for fibro myalgia[99], Parkinson's[100] and physical[101,4] and cognitive fatigue[102]. Assessments used included Stroop, Finger Tap Test, FTT, Trail Making A, Trail Making B, paced serial addition test, PVSAT, memory, jump height.

Table 5-1 Assessment battery

Assessment	Bio-psycho-central performance	Reference
Finger Tap Test	Neuro muscular fatigue	[101]
Stroop	Cognitive flexibility and selective attention	[130]
PVSAT	processing speed, attention, working memory	[135]
Trail Making A and B	Motor and executive impairment	[132]
Corsi Block test	Spatial memory, Working memory	[147,148]
Vertical Jump	Neuromuscular fatigue	[138]
Rating of Perceived Exertion	Perceived level of exertion	[4,221,230]

Data Preparation

The participant wore a chest mounted BioHarness (Medtronic, Minnesota, USA)[204,205] for acceleration data (100 Hz, vertical x-axis, sagittal z-axis, lateral y-axis) and electrocardiogram (ECG) (250 Hz) and a Garmin Forerunner GPS (Garmin) wrist watch (1 Hz, horizontal accuracy 6 m) to assist with labelling and location.

The trail was divided into twenty three sections separated by waypoints defined by a change in terrain surface, slope or obstacle. Terrain descriptors were validated against video (GoPro Hero Session, 1080p 25 fps). Slope was determined from a mean of GPS altitude measurements at each waypoint. Waypoint location was determined from Google maps to an accuracy of 10 cm. Time at a waypoint was determined when the subject was closest. Walk and Run activity labels were defined by cadence from vertical axis accelerometery zero crossings ($100 < \text{Walk} < 150 < \text{Run}$ steps per minute).

Identification of crossing obstacles was based on geographic location and manual observation of the acceleration waveforms. Time resolution for labelling was one second.

Convolutional Neural Network

Figure 5-2 shows the multi-channel 1-D Convolutional Neural Network (CNN) that was selected to allow learning on separate channels and cross correlation into a single regression output value. The training label was FTT up sampled to 250 Hz. Data were split by activity type and segmented by input window length. The initial model width for all hidden layers was set at 256 which was approximately one second of data. The model implemented the Adam optimiser and mean absolute error (MAE) as the error term during training. Randomized train test split ratio was 0.33.

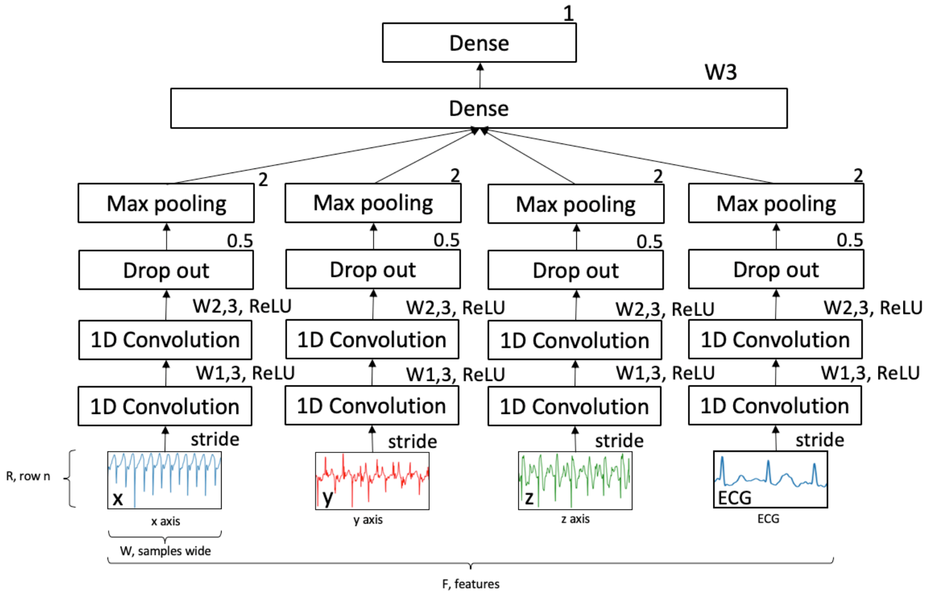


Figure 5-2 Structure of CNN

Tuning the window size for each activity type was performed (64, 128, 256, 512). The lowest MAE activity was selected for further model optimisation of hidden layer widths. Optimisation was performed separately for three datasets: acceleration; ECG; and combined acceleration and ECG. The final model for comparison was selected for lowest MAE. Performance was assessed using the mean absolute difference (MAE_{200}), and range of absolute difference (RAE_{200}), between the label values and the average of

200 predictions. RAE was of interest as it indicated the largest error possible when the trained model was used to predict a fatigue value.

Statistics

Linear regression (Pearson correlation R^2) was performed on each performance test to assess sensitivity of the protocol. The performance test results were normalised across the protocol and linearly interpolated to give a long-term linear fit (LTLF). The tests with highest R^2 for LTLF were up-sampled to 250 Hz using inter-test interpolation (ITI) as ITI is includes short term variations which could represent short term fatigue and recovery. LTLF more representative of long-term fatigue but is only possible with a research protocol designed with constant a load over time. ITI is needed for random field predictions where no assumptions can be made about overall loads.

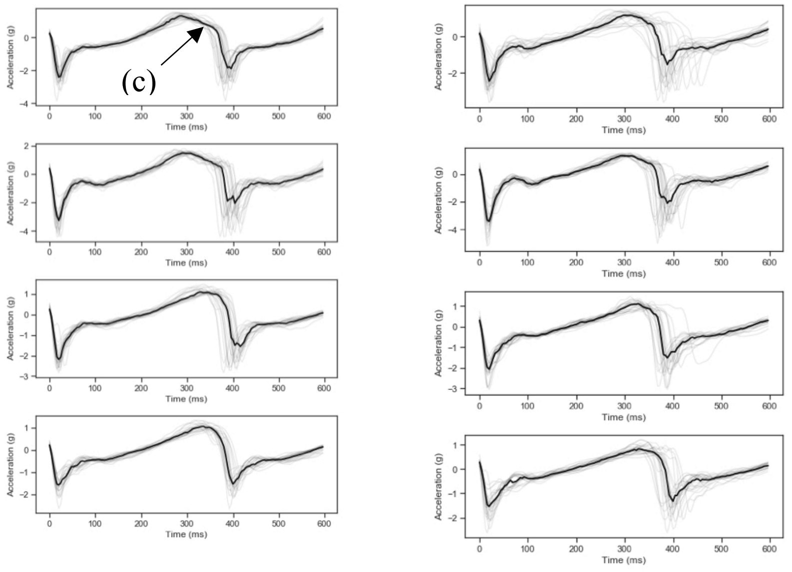
Time series data were normalised using feature scaling via Equation (5-1 in preparation for training the CNN. ECG data were base line corrected. All accelerometer axis (x,y,z) and ECG data were transformed into an array (D,W,F) with D rows, W window width, and F number of features.

$$x_{\mathrm{new}} = \frac{x - x_{\mathrm{min}}}{x_{\mathrm{max}} - x_{\mathrm{min}}} \qquad (5\text{-}1)$$

Results

The participant voluntarily ceased the protocol at 11 hours (2,200m vertical climb, 41.8km) due to perceived exhaustion.

Figure 5-3 shows the representative input to the CNN of the gait waveforms of vertical acceleration on tarseal and dirt at different fatigue levels. Each plot is 50 steps triggered at zero g and plotted with the median waveform in a thick black line. Inter-step variation in acceleration and morphology can be observed between surfaces (a) tarseal and (b) dirt. The changes in waveform shape between surfaces was likely due to surface hardness and variations in surface texture uniformity. Across the protocol variation was likely due to fatigue reducing peak forces and subsequent gait adaption as seen on the plots at point (c).



(a) (b)

Figure 5-3 Acceleration for activity "Run Down" over protocol time (0,3,8 and 10 hours) for surfaces (a) tarseal and (b) dirt and (c) feature of interest over fatigue.

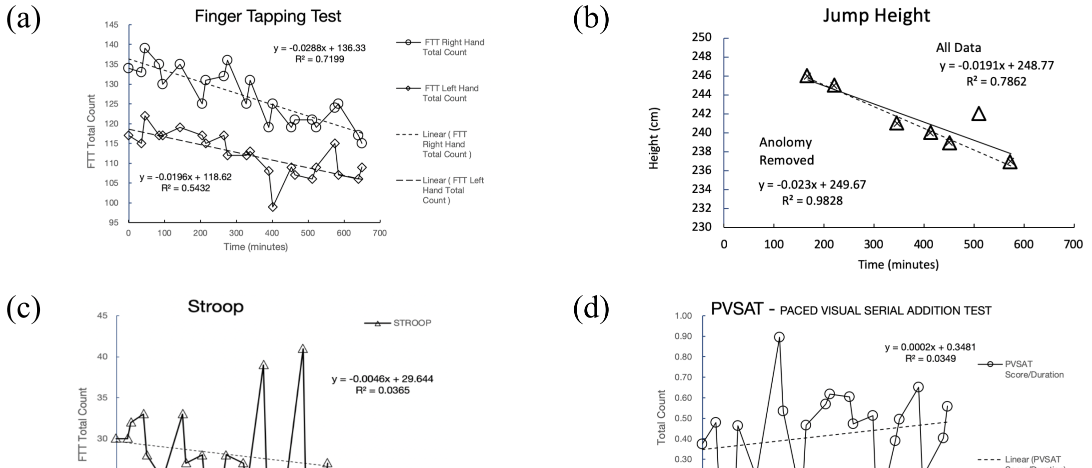


Figure 5-4 Performance Tests (a) FTT (b) Jump Test (c) Stroop (d) PVSAT

A subset of performance tests (FTT, Jump test, Stroop, PVSAT) completed in the protocol are shown in Figure 5-4.Jump and FTT-right-hand were most sensitive to the fatigue protocol. Figure 5-4 (a) shows FTT-right-hand and the slower non dominant left hand with separate linear regression lines. Inter-test variation was observed between physical and cognitive tests with an overall trend having a negative slope showing performance was decreasing over time. Correlation results for all tests are shown in Figure 5-4. Jump height shown in Figure 5-4 (b) was performed after each physical load period and showed high correlation (R^2 0.78) with the protocol. Stroop shown in Figure 5-4 (c) had two outliers and showed moderate correlation (R^2 0.5) with the outliers removed. PVSAT shown in Figure 5-4 (d) was not correlated with the protocol load. Trail making A (R^2 0.29) and spatial memory (R^2 0.28) where somewhat correlated to post cognitive load. Trail making B (R^2 0.22) was somewhat correlated to post physical load.

Table 5-2 Performance test sensitivity, R^2, to the protocol load

Test	All Tests	Post Physical Load	Post Cognitive Load
Jump	0.78	-	-
Finger Tap Test			
Dominant Hand	0.72	0.76	0.67
Non Dominant Hand	0.54	0.51	0.60
Stroop (with outliers)	0.04	0.003	0.36
Stroop (no outliers)	0.49	0.37	0.36
PVSAT	0.03	0.11	0.02
Trail Making A	0.19	0.04	0.29
Trail Making B	0.001	0.22	0.05
Spatial Memory	0.00	0.00	0.30

A training result is shown for a single activity ‘run down’ in Figure 5-5 for data window 128, epoch 100, individual predictions (light grey) and rolling average of 200 predictions (black). The label for FTT (red) is inter-test linear interpolation with discontinuities between time periods due to concatenation. The difference between the label and average predictions in Figure 5 shows the range from -0.1 to 0.075 with the largest error at the start of the protocol.

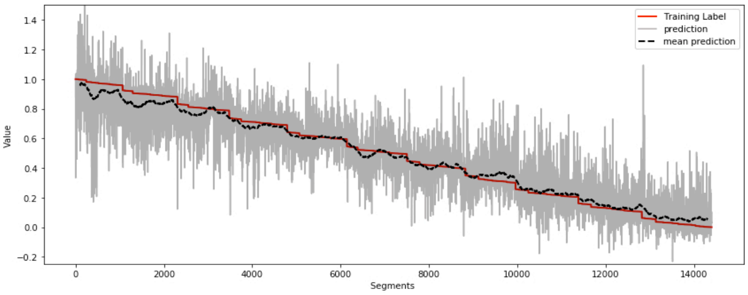
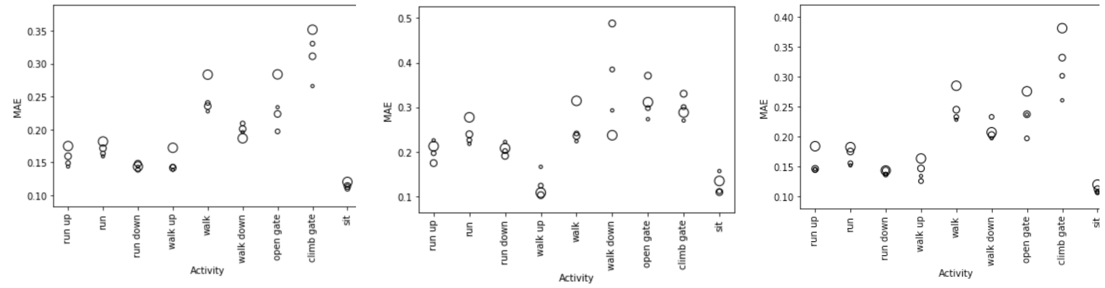


Figure 5-5 Training Results (BLACK) for CNN with activity 'run down' with training label (RED) and individual predictions (GREY)

A total of 108 experiments were performed to test which input data width and activity type gave the best MAE. Initially a fixed CNN topology was used (Epoch 50, Batch 256, layer 1 filter 256, layer 2 filter 256, dense layer 128, overlap = 0). Three data group results were compared for; acceleration, ECG and combined acceleration with ECG. These three conditions were tested for each activity type ('run', 'walk', etc) over four data window widths (64,128,256 and 512). The results for these experiments are shown in Figure 5-6 by activity where circle diameter is data window width. Minimum MAE was at 'walk up' (window width 256, MAE 0.105,samples 1534500, windows 5994) and 'sit' (window width 256, MAE 0.116, samples 2662750, windows 10401). However, sit was not included as it took place in the lab for cognitive testing. Samples were more numerous for 'run down' (window width 256, MAE 0.181, samples 1843749, windows 7202) and still gave a larger minimum MAE. This indicates total sample count is not the main influence on MAE however the activity with considerably lower samples did show larger MAE values, 'walk down' (stride 512, MAE 0.309, samples 20000, windows 78).



(a) **(b)** **(c)**

Figure 5-6 Training Loss, MAE, for (a) Accelerometer, (b) ECG and (c) combined ECG and accelerometer, epoch = 100, circle radius set by data window width (64, 128, 256, 512)

Further experiments were performed for acceleration and ECG with ‘walk up’ to optimise the CNN model hyperparameters, various widths of the first two convolutional layers and the dense layer. The lowest MAE was found to be the following model; Conv1D 128, Conv1D 128, max_pooling, flatten, dense 128, dense 1.

Table 5-3 shows the total samples per activity and results for MAE and RAE with the training labels using two methods, linear fit and inter-test interpolation, window width 128, epoch 100, batch size 256, rolling window average of 200 predictions. There was no result for activity of ‘walk-down’ as the total samples divided by the window width of 128 was 156 which was less than the rolling average of 200 predictions. Activity ‘Walk Up’ gave the lowest MAE for both linear interpolation and inter-test interpolation of label data. Activity ‘Run Down’ gave the lowest Range of errors, indicating it may be a better activity for field prediction.

Table 5-3 Results for linear fit and inter-test interpolated labels.

Activity	Data (250 Hz)	Linear Fit MAE_{200}	Linear Fit RAE_{200}	Inter-test Interpolation MAE_{200}	Inter-test Interpolation RAE_{200}
Run Up	1,019,002	0.145	0.225	0.134	0.240
Run	732,501	0.151	0.238	0.156	0.232
Run Down	1,843,749	0.130	0.289	0.133	0.167
Walk Up	1,534,500	0.136	0.303	0.125	0.411
Walk	299,997	0.238	0.683	0.235	0.726
Walk Down	20,000	0.219	-	0.239	-
Open Gate	56,750	0.195	0.338	0.199	0.316
Climb Gate	65,249	0.327	0.422	0.313	0.389

Discussion

A protocol for cognitive and physical fatigue was performed in the field with voluntary activity selection over various terrain slopes and surfaces. Jump and FTT-dominant-hand were most sensitive to the protocol. FTT-non-dominant-hand and Stroop were moderately sensitive. FTT was most sensitive and biomechanically non-specific as the legs were exposed to physical load and the arms-hand-fingers were tested for neuromuscular performance. It is likely Stroop would be more sensitive if the protocol included sleep deprivation. Spatial Memory was mildly correlated to the cognitive load.

The experiment showed that a field protocol of cognitive and physical load in excess of a critical power will cause failure and modulate standard objective measures of cognitive and physical performance. Mental and physical fatigue led to earlier than anticipated termination of the protocol which aligned with previous studies[4,14].

The use of a machine learning model was required due to the variable gait waveform morphology. The results for acceleration, ECG and combined acceleration and ECG are shown in Figure 5-6 across various stride lengths from 64 to 512 samples. While the activity ‘sit’ had low MAE showing how a controlled environment could give good results, our work aimed to determine if it was possible in an uncontrolled field based environment. Activity ‘walk up’ had low MAE for both inter-test interpolation and long-term linear fit. ‘Run down’ had the lowest RAE. It is recommended that RAE is used as this represents the results you would get when using the model in the future as a prediction model.

This experiment showed a single sensor could be used in conjunction with a CNN model to give accurate results of cognitive and physical fatigue equivalent to gold standard objective tests; that is FTT and Vertical Jump Test. Best results were obtained when model training was specific to activities such as ‘run down’ and ‘walk up’. MAE and RAE performed well for a rolling window of 200 continuous predictions of 102 s. This intuitively makes sense that any one step in a persons’ gait may be influenced by objects, surface and other distractions and it is best to use multiple steps of a persons’ gait to determine a fatigue result. Winter[231] showed the cadence in steps per minute on a uniform surface varied from 84.7 ±10.4 for slow to 121.6 ±5.3 for fast.

The input window size of the CNN model has an optimum size: Too small does not allow a full gait or ECG waveform to be analysed; Too large significantly reduces the number of training samples.

Tests that had highest sensitivity to the protocol, and indicated a central fatigue component, were the jump test (high physical load on the legs), and the FTT (utilised hand digits which were not significantly utilised during running). Cognitive tests were less sensitive to the protocol indicating there may have been a mismatch between cognitive and physical loads.

The effectiveness of the protocol was encouraging as it provided proof of concept for translational research to be undertaken in outdoor environments. Future work could examine how team workload and tactical decision making can be adjusted for cognitive and physical fatigue in real time with no additional data entry for soldiers on multiday missions. Recovery during training missions could be assessed without researchers being present. Adventure sports people could gain insight into their cognitive and physical fatigue enabling informed training plans. Work rest cycles could be adjusted and critical tactical and navigation decisions chosen based on periods of highest cognitive performance.

This feasibility study researched approaches of protocol design, error sources, calibration techniques, data collection, validation, labelling and data processing. Given the lessons learnt, data gathering and processing needs to be more automated to reduce the high processing load that occurred for the one participant in this study. Further work is needed to test inter-subject variability to the protocol.

Limitations

Limitations in validating the experimental objective include a linear protocol and the limited amount of comparison tests, however, this is a natural limitation in the field of cognitive assessments in the field. A long-term linear fit was appropriate for this protocol as the repetitive load could be assumed constant over the longer term time frame. A random field assessment with no defined load protocol would require training using inter-test interpolation to allow for stochastic loads and recovery cycles. A constant long-term load was required to fit a machine learning model. Future work could compare the results in a long-term non periodic protocol.

Limitations of this test were the duration and the use of a single participant to initially prove the feasibility of the protocol and approach. Further research is required into increasing the duration of the protocol, possibly by reducing the hourly physical load. Additional studies over longer periods are required to generate cognitive fatigue that includes sleep deprivation. The test battery should include assessments immediately after large vertical assents to gather insight into short-term recovery. The addition of cognitive loads and assessment significantly affected the rate of perceived exertion. Future protocols should half the physical load to lengthen the time to failure. Additionally, this method requires more participants to compare inter-person sensitivity and variability.

Conclusion

This paper showed that a single wearable sensor could be used in conjunction with a neural network model to determine cognitive and physical fatigue without performance tests being required during an operation in an outside unstructured environment. This research has potential to increase safety and operational performance in high risk environments by indicating the possibility of replacing traditional performance tests with a single wearable device. This work is novel to the authors knowledge in developing a field based protocol for human performance with no direct supervision and modulation from ground surface, slope, fatigue and task motivation.

6. DISCUSSION

The following chapter provides an in-depth discussion of this research and highlights the contributions this work provides to the wider field of sports and exercise science. The insights gained from the field research, limitations and potential further work are also discussed.

Definition of Fatigue

The definition of fatigue proved to be complex and often argued. The study of fatigue is multi-disciplined including the fields of: physiology, biomechanics, neurology, clinical psychology, sensors and mathematical models and deep learning for time series modelling. Research into fatigue has been ongoing for over a century and a summary understanding is instructional to understand the complexity and how a reductionist approach in the scientific method may give rise to certain mechanistic understanding but has delayed fatigue research from understanding how complex fatigue is and what is required to pursue further research. This desire to maintain a simple model in the presence of variable experimental results may be due to insufficient access to tools that can compute complex situations. It is possible that artificial intelligence provides a research tool which is necessary to continue our understanding of fatigue.

Fatigue was first studied from a physiological perspective as muscle energetics in the late 1800's by Mosso[12]. In 1924 Hill[16] developed his theory where maximal oxygen uptake (VO2) limits maximal exercise and later the theory that lactic acid was a limiting factor in exercise due accumulation in muscles from insufficient oxygen delivery. This theory held for quite some time as it explained certain failure mechanisms, however experimental results varied indicating there were other unknown factors limiting performance.

The introduction of neurological perspectives of fatigue, often referred to as central drive, with its effect on pacing was developed by Melzack[20] in 1965 and the rating of perceived exertion (RPE) scale which is still in use developed by Borg 1969. The RPE scale is interesting as Borgs' theory introduced the concept that perceived exertion could vary compared to actual performance and that the perception of exertion was the limiting factor not the actual physiology in the body.

The concept of motivation and goal versus cost was introduced by Klinger[22] in 1975 where a person made a decision to continue at a given pace or changed goals in order to survive or be more comfortable. The concept of a plateau with various physiological couplings continued for several decades with work by Wasserman[17] with anaerobic threshold and protocols for assessing these plateau such as Conconi's[19] ramped effort test. The Conconi protocol is interesting as it is often performed on a treadmill and can involve researchers encouraging subjects to greater effort which is removing the limiting factor of perceived exertion and motivation that would be found in settings out of the laboratory.

In 1992 Enoka[25] published a paper addressing the psychology of muscle fatigue where a cross over from central fatigue to the psychology of fatigue had begun where one of his evaluations included task dependency. Noakes[11] in his 1997 paper titled 'challenging beliefs' described examples that he argues disproves the physiological understanding that exercise is limited by muscular-energetic or cardio-respiratory or central drive mechanisms. He argued any limitation is due to an anticipatory decision to preserve the body's internal systems. The central governor theory is possibly the most argued in the literature, however there are other similar tripartite theories that include physiology, neurology and psychology [13,232,233].

The tripartite model continues to be used and is now made more complex as the psychological aspects require research to incorporate real world goals, motivations and distractions. Enoka[25,234,235] is well published on fatigue and the effects of the brain on muscular performance. In his 2016 paper he concluded that fatigue research will only be further understood if it leaves the laboratory and is researched in the field with all the complexity that entails. It is interesting to note that a lot of fatigue research is based on end points of failure, however most field activities are not performed to failure and it is more useful to measure performance and fatigue on a continuous basis.

Based on this research, it is my opinion that fatigue is not due to any one failure mechanism but a cascade of resources and perceptions that will change case to case and person to person. The reality of fatigue is that it is influenced by the current state of the person, their environment and goals. Simplified protocols will always risk leaving out an important factor, which may help understanding of an isolated part of the whole but will not be repeatable or representative of the individual in undertaking an activity in a real environment. This thesis takes this learning to investigate if it is possible to simulate a complex field environment and also gather laboratory accurate data as validation. The field studies were designed to be realistic with all the variation and challenges normally incurred and investigate if validation protocols could be adapted to be accurate and logistically possible.

Effect of Environment on Fatigue

Moving experiments into the field required a survey of previously research environments. These were found to include sports, ultra-events, industrial, military, driving and clinical. Field research was instructive to give further understanding of previously validated protocols, sensors, assessments and analytical approaches. Each environment had different limitations resulting is a unique protocol of loads, sensors and assessments.

Driver fatigue is instructional for the definition of fatigue where the Volkswagen research team[2] defined it as the "speed to respond to an unexpected event" with underlying cognitive mechanisms including scene scanning, scene cognition, spatial modelling, anticipation, neuro muscular signalling and peripheral force. This definition of driver fatigue offers a definition of fatigue as a reduction in ideal performance and also a guide to determine the various mechanisms required to perform that task and how they may become depleted during normal task operations. Interestingly the Volkswagen team noted electrooculography (EOG) could detect spatial scanning but not spatial cognition such that the eyes continued to scan the scene but the brain stopped processing the information. This was my first insight into the complexity of fatigue and how the bodies overlapping redundant systems require assessments to determine performance measures well before exercise cessation or a maximal plateau is reached. Understanding which mechanisms of physiology, biomechanics and neurology are relevant to performing a mission is possibly the most challenging in fatigue research. Borghini[114] studied pilots and drivers with EOG and EEG. Data collection for driving is often performed in quiet laboratory rooms with no distractions. This allows various response measures to be collected including EOG, EEG. However the lack of distractions and head mounted equipment did not lend themselves to energetic outside environments which was the area of study for this thesis. Furthermore any model needed to be accurate in the presence of distractions to be generalisable in the field.

Industrial applications have led to the development of assessment batteries which measure several criteria relevant to a given mission. For pilots NASA has developed the multi attribute test battery (MATB)[98] which assesses multiple cognitive and executive decision making processes relevant for an aircraft pilot. The MATB was re-developed in Python code as part of this research and used as a cognitive load requiring multiple task types to be performed in the second field study.

Clinical applications of fatigue are very useful for comparison to athletic activity as they can offer insights into particular performance reduction and possible data collection methods. Parkinson's is a neurological disease and has shown significant effect on gait[4,236–238]. This led to the research considering cognition affected gait as a likely dataset to be sensitive to a cognitively fatiguing protocol.

Sports applications on fatigue can include several categories: maximal voluntary contractions, endurance events with sensors for post event analysis and fatiguing protocols to determine measurable effects. Buckley[171] used three IMU sensors on runners and a measurement protocol using a 400m track running at their 5km self-pace cadence. A beep test was used as a fatiguing event in-between the 400m runs. A machine learning model was used as a binary classifier for fatigued or not fatigued. This experiment shows a well thought out protocol to induce fatigue with three sensors and machine learning which is instructional for the work undertaken in this thesis. The limitations of this experiment for remote long duration events include no cognitive fatigue, a smooth flat surface, three IMU sensors and a limited goal that may have had no effect on motivation. Clermont[239] recently published a paper where the fatiguing event was a marathon with an IMU at centre of mass to characterise gait. Assessments were carried out pre event and post event at two and seven days using a 200m oval track surface. This paper was a good indication of how a single IMU could be used to characterise fatigue effects on running biomechanics including neuromuscular fatigue and musculoskeletal pain. Both papers indicate a method of induced fatigue and measurement of results that could be extended to the field using IMU sensors. Uneven terrain and cognitive fatigue were not covered in the papers which are gaps for research.

The field studies in this thesis addressed the gaps in previous research by conducting experiments in real outdoor scenarios to ensure all environmental effects on the participants were present. Additionally the research investigated what protocols would enable the results to be validated accurately. This resulted in new questions including what sensors would be practical, what assessments could be used for validation and how to determine the actual physical and cognitive loads.

Fatigue Assessments

Assessments of fatigue is required for validation of results and training machine learning models. Clinical examinations have been validated that measure various aspects of cognitive performance listed in Table 5-1. These clinical assessments[240] normally take hours and a subject may be assessed no more than several times a year which limits any learning effects. In this research assessments where shortened to minutes and taken hourly. To reduce training effects the assessments were practiced several days before data collection. The protocol used in both experiments for this thesis used multiple assessments to determine which test was sensitivity to the cognitive and physical loads and address possible reduced sensitivity from reduced assessment duration. Assessments were chosen that have been published in similar analogous studies or previously indicated for the expected reduction in performance for related neurological symptoms, such as Parkinson's disease. Part of this research included the sensitivity of assessments when the test time was shortened and the subject took the test multiple times per day. The first field trial showed the tests were sensitive but one per day was not sufficient for training machine learning models. Additionally using a separate app for each type of test was cumbersome and would not scale. The second field trial used custom developed apps for both assessments and cognitive load which enabled hourly testing. The custom apps resulted in consistent loads and were efficient with complete compliance. Custom applications for each field protocol is recommended for both the subjects under test and the researcher who is often fatigued and error prone due to the length of the protocol. Some tests were not as sensitive as expected, such as Stroop, however further investigation indicated this may be due to early cessation of the protocol and sleep deprivation of more than 5 hours is required before Stroop is affected.

Single Sensor Selection

The aim of this research is to determine if a sensor and ML model can replace manual fatigue assessments in the field. Memory, power limitations, size and weight are all limiting factors for what sensors are possible and need to be taken into account with previously published[4,236–238,114,171,239] work on what physiological and biomechanical signals have shown sensitivity to fatigue. Sensors such as EOG and EEG are not appropriate in the field due to cables, size and limiting a person's movement.

Verlinden[8] showed gait and cognition were related. Buckley[171] and Clermont[239] have shown gait is affected by physical fatigue. Gait is a useful measure for field studies as it can be measured with an inertial measurement unit (IMU) which can be small, unobtrusive and battery powered. Electrocardiogram (ECG) was selected as a measure of cardiorespiratory system fatigue. The chest strap used for this research gave good results in the second field study which was based on trail running. The use of a chest strap for sailing resulted in low compliance due to comfort. Future work investigating wrist sensors is seen as the most compliant option in the field as a watch provides additional value to the subject by displaying their personal data , time of day and navigation. Body shape and fit is less of a concern for wrist devices over chest devices, however adequate coupling of optical plethysmography can be a challenge resulting in lost data. Several very small sensors at the waist and shoe may also be possible, however data management is a significantly large task and needs to be considered when multiple subjects over multiple days is investigated. While this research investigated a single sensor it did include an additional GPS watch for labelling. It would be worth investigating a GPS watch with IMU and optical heart rate and possibly an IMU attached at the waist to remove the need for a chest strap. One of the least expected issues with most commercial sensors is the lack of access to raw sensor data. This is a significant limitation to sensor choice when using deep learning. The BioHarness was the only device found with accessible ECG. The Polar H10 does have beat to beat which may be sufficient for some fatigue studies, however ECG has the potential to show metabolic deficiencies that may be of interest and detectable with deep learning.

Artificial Intelligence Models

A subfield of artificial intelligence is machine learning (ML). ML can be; (i) unsupervised where data is grouped using algorithms that attribute certain values to multi axis similarity or (ii) supervised where data is labelled and trained. The output of an ML model can be the prediction of a most likely class, called a classifier or the prediction of a value, called regression. Traditional machine learning process statistical features, requiring the researcher to have knowledge of the dataset features that will change over the solution space. Additionally this approach is limited to known features that are calculatable. Deep learning operates on the data directly which removes the step of feature extraction, but may require more data. All ML models require representative datasets to enable good results for future unseen datasets, this is called generalisability. ML research follows a process of data collection, model design and model training. Improvement in model performance can be attained through either improving the model or collecting new data to cover more use cases. This thesis investigates if an ML model performs equivalently on field datasets that have not been collected previously, where the data is significantly more variable and includes noise from modulating factors such as: fatigue, variable terrain, variable slope and obstacles. Previously published datasets have been on smooth surfaces in predictable environments, such as a laboratory or running track which results in very uniform data. Additionally due to the requirements of remote multiday events a single sensor will be used to collect data, where previously published models have used multiple IMU sensors[171], or data from sensors that are not practical in field such as Zhang[241] EOG, ECG, EEG + ANN. Multiple sensors have advantages such as IMU on different parts of the body, for example detecting sitting and standing is made easier with the addition of a thigh IMU along with a IMU attached to the thorax. A single IMU is more practical in the field but may suffer from detecting certain activities.

Deep learning models have an input layer and several hidden layers. The number of neurons in each layer is called the width. The type of neuron is characterised by its activation function and the weights from all neurons in previous layers. These activation functions perform differently and are part of the ML model design process. The output of the model can have one or many neurons and be configured for classification or regression by changing the activation function. The model architecture and weights define a trained model. Deep learning models are trained on data using a method called back propagation[242]. First forward propagation is used on the data and then back propagation is used to adjust weight values to minimise overall errors using an algorithm called stochastic gradient decent.

Deep learning has two standard models: recurrent neural networks (RNN) and convolutional neural networks, (CNN). RNN have traditional been used for time series due to their recurrent nature however CNN have become more common in recent times[162], in a recent review by Cust[197] the ratio was 12 to 1 CNN to RNN for activity recognition in sporting events. CNN models have been used on time series data[161] for emotion detection[167] and mental fatigue[168]. A CNN uses multiple types of convolutional filter to automatically detect features of interest. Hidden convolutional layers determine more complex features. Drop out layers reduce connections during training as this has shown to generate models that are more generalisable[243,244] to new data.

Previous studies have not used datasets with terrain and fatigue variation. Human activity recognition has been performed in controlled environments with repeated tasks over the same track or equipment, such as "walk down stairs". While 20 people may have walked down the stairs, it was the same set of stairs, with no obstacles to avoid, trip hazards or varying degrees of fatigue. The person knew they were about to walk down the stairs and could plan the activity. This thesis investigated if it was possible to collect data where the person was free to change their activity second to second based on the environment or level of fatigue and without knowing the terrain immediately ahead. For instance they may choose to walk because they enter a section of long grass or running down hill on mud making traction difficult or stop running up a hill due to acute musculoenergetic or cardiorespiratory exhaustion. Additionally, obstacles such as rivers and fences required changes to repetitive activity for obstacle crossing. A CNN model was chosen for this research due to its reported robustness for data variation and potential for battery powered devices in the future as the CNN model can operate with sparse memory requirements.

A CNN model was chosen for this research due to its reported robustness for data variation and potential for battery powered devices in the future as the CNN model can operate with sparse memory requirements. Traditional ML models such as support vector machines, k-nearest neighbours and random forest have been used in laboratory settings for human activity recognition. These measures where not used in favour of a DL model due to the large variation in the time series data due to random texture of a mountain trail. The protocol allowed simple measures such as lap time to be used as a fatigue measure. However this would be constrained to a lap based protocol and not a single large lap course in the future, so was not considered for the research question.

The CNN model performed well for HAR after the activities were grouped by major activity, e.g. running up, running flat, running down. The data labelling took significantly longer than planned as the voluntary nature of the activities led to unstructured periods of various activity types. Labelling accuracy quickly became a limiting factor to ML model accuracy and this required each participant step to be analysed. For example, when a person approaches a fence they may slow to a walk for several steps, stand, climb, stand, walk and then run again. Decisions needed to be made by an expert labeller on each of these stages of activity if the true accuracy of the model was to be understood. The results of this research would indicate that periodic longer term exercise provides good results and time saving approaches for labelling may be possible such as only including running or walking with ten continuous steps. This would enable a significant amount of automated labelling at the expense of removing some data from the training set. Now this research has been conducted it would be recommended to ignore the small unique activity variations and increase the number of subjects.

The CNN for fatigue performed well once HAR was modelled and used as an input into the second model. This research has found that certain activities are strong indicators of fatigue, such as “run down” or “walk up” and this requires a cascaded model approach.

The CNN models were quick to train on a Mac Book Pro taking up to a day to train different scenarios. The use of a super computer was investigated, however the additional development work was deemed greater than existing methods on a local machine. If a super computer was used then other scenarios could be investigated such as wider input layers to the model as over 1000 neurons were not possible on the Mac Book. It is questionable if this is a useful avenue of investigation as it will be unlikely to result in a trained model that can run on a battery powered device in the field.

The use of the low pass filter after the CNN model is novel to this researcher and came about from the individual predictions having a large range. The issues this uncovered lead to unique learning that a person's single step is not indicative of fatigue and several strides are required to make a prediction. The number of strides required increased as terrain variation was added which made sense as the terrain changed the stride pattern as the person navigated small obstacles such as logs, rocks and trenches.

Field Study 1

Field study 1 for this thesis was to investigate a fatigue protocol during an offshore sailing voyage over 12 days with 4 person crew. The protocol had Ethics approval AUT 17/353. Cognitive assessments were performed via separate iPhone apps and a collection of Likert scales were collected along with a BioHarness wearable sensor with IMU and ECG. The results from this field study lead to several important learnings: (i) a single software application is required for each subject to reduce redundant data entry, (ii) a known and repeatable physical and cognitive load is required for this enable validated field, (iii) daily assessments are insufficient to asses variation and train a machine learning model (iv) short duration cognitive assessments are sensitive. This field study also had less than expected compliance to the protocol due to distractions, monotony and lack of sleep. A major contribution of this work was to derive a compliance formula that addresses motivation and cost of performing a task. This allows protocols in future to be planned for maximum compliance.

Compliance

This research has led to a compliance formula which allows the deconstruction of a protocol to maximise compliance. Compliance is important as there is a trade-off for cognitive assessments between sensitivity and the frequency of testing to capture short cycles of fatigue and for labelling of machine learning models.

Field Study 2

The field study 2 took lessons from field study 1 and previous papers including Verlinden[8], Buckley[171] and Clermont[239]. The key features of the protocol included repeatable loads, performed hourly with a single software application which was custom designed with cognitive assessments and Likert questionnaires, shown in Table 5-1. Ethics approval was AUT 18/412.

The second field experiment introduced a challenge to the question of whether a field trial can combine extreme outdoor environments with a challenging goal while also providing laboratory quality validation. This required a new protocol design using an outside trail with; (i) variation in terrain (slope and surface), obstacles to create variation; (ii) a sufficient distance and vertical climb to act as a fatiguing load without early cessation of the protocol; (iii) an area where a computer could be used for cognitive tests and cognitive load; (iv) a protocol that would last 24 hours to invoke sleep deprived cognitive fatigue symptoms; and (v) a goal of 100km in 24 hours with 15,000 feet vertical climb in order to invoke psychological goal versus resource action crisis[23].

Various logistical steps were required that included selecting a trail, physical training, cognitive assessment practice to reduce training effects, development of cognitive load software based on MATB[179] written in Python code, iPhone app based on Apple Research kit and additional Likert tests and calibrate the trail. Initial planning with a trail distance of 3.8km 200m vertical climb was selected. It had various obstacles including three farm gates to climb, two stream crossings, stairs, trees to be avoided, trails to be navigated, surfaces including tarseal, gravel, dirt, clay, mud, long grass and a vertical climb that could not be run to the top requiring a transition to walking at anaerobic threshold. The slope varied sufficiently to generate different gait for steep up hill and downhill. Training began to become fit enough to complete 100km in 24 hours as the stated goal.

Software development for cognitive load was based on the NASA MATB test implementation based on the original paper[179]. MATB is designed for pilots and has related tasks including attending unexpected alerts (ocular scanning and respond), occasional target tracking tasks (ocular scanning and spatial planning) , audio response to radio calls with identification and requests to change radio channels (audio scanning, processing and response), fuel management with faulty pumps (ocular scanning, fault finding, strategic thinking with long term planning). The MATB application ran for a fixed period of ten minutes then automatically stopped to ensure an exact cognitive load. The second app was an iPhone app built using components from the Apple Research Kit. The software was designed by myself and coded by an experienced app developer. Based on learnings from field study 1, the app was assigned to one person which meant name and age were entered only once with an email address added for results. A single button press was required to start the protocol and results were automatically emailed at the end of each test to avoid data loss.

Calibration of the field trail required satellite imagery to assess waypoints within six decimal places for latitude and longitude accuracy of 10cm. Terrain, surface and obstacles were recorded for each section between waypoints. A topographic map and GPS was used for altitude and slope. The course was videoed during walking with two people confirming all records and video validation. This confirmed a completely remote assessment is possible for future studies.

This field experiment was new compared to previously published work and investigated a protocol that could measure fatigue from various loads including: physical, psychological and cognitive. It challenged the concept of field research with laboratory grade validation and if hourly labelling of fatigue was sufficient to train an ML model. The terrain and fatigue variation resulted in data that was extremely noisy and would challenge an expert to derive insight and hence was interesting to see if deep learning could perform. The research goals were broken down into two experiments, firstly can activity be classified and then can fatigue be predicted.

Field Study 2, Experiment 1: Human activity recognition (HAR)

The first experiment was limited to HAR as it was unclear if a machine learning model would perform with the level of terrain variation, gait variation and fatigue. Figure 4-1 shows increased variation between a smooth surface and rougher terrain. It would be quite a challenge to tune statistical analysis models to derive features in a traditional manner from this data.

It became quickly apparent that data labelling required some automation to gain any context of the 100Hz IMU data. Firstly the waypoints of the course were aligned with surface type (concrete, gravel, grass, dirt) and slope (up, flat, down) and any obstacles (gait, river). An important learning was that manual labelling was only possible with this context and calculated cadence from the acceleration data. This took an astoundingly long time over several months and confirmed other researchers reports that have noted the amount of time to wrangle data is ten times the effort of performing the ML model design and training. Collection and labelling methods for the dataset is a significant contribution of this research.

The results shown in Table 4-3 showed the amount of data collected for the trail run was approximately three times larger than the well-known WISDM[208] dataset which is used in many HAR publications. While our dataset had a single participant we have variation in fatigue, terrain and significant slope variation generating different gait. The WISDM dataset had 22 subjects but no variation in terrain or fatigue, all surfaces were even and predictable with no obstacles. It is interesting to note that the subjects here could view their task before performing it and had an understanding of surface regularity and friction, all factors that do not exist in the field.

The accuracy of the model was 97.5% for all five activities of climb gate, lay, sit, walk, run including run up, flat and down. The least accurate activity was the non-periodic "climb gate" with an accuracy of 80.2%. These results compare to the indoor WISDM dataset of 97.7%.

ECG data contributed to the accuracy for activity 'walk up' and were less significant for 'run down'. This result intuitively makes sense as 'walk up' is more physiological demanding. Of note the CNN window was tested between 64 and 256 samples (0.25 to 1 second approximately. This would preclude heart rate, typically average of 15 seconds, and heart rate variability, typically SDNN300 using 300 beats. Future work could consider adding these long term measures to the data.

This result was as accurate as laboratory results and shows deep learning can automatically determine significant features. As with all ML applications further work is required to determine if these results are generalisable to more participants and in various field environments.

Field Study 2, Experiment 2: Fatigue prediction in the field using a single sensor and an AI model

With the positive results and a ML training protocol from experiment 1, the next step was to model fatigue. Firstly the assessments were compared to determine which was most sensitive to the protocol load with the shortened test duration.

The results over time for each assessment are shown in Figure 5-4. Finger tap test (FTT) ($R^2 = 0.72$) test and jump test ($R^2 = 0.79$) were most sensitive. The FTT was used for training as it had more data points and is representative of both neurological and muscular fatigue. It is interesting to note that the FTT was influenced from physical loads in the legs when its measures the response of a finger. This is validation of central fatigue.

It is recommended that all tests are kept in this protocol as different people may vary in response and with reduced physical load the protocol would last 24 hours which may increase the Stroop effect.

The results for predictions from the CNN model for every window are shown in Figure 5-5. This figure is important for several reasons. Firstly the grey line shows the range of prediction accuracy. MAE is used as the feedback during training and is often reported as the figure of merit in ML papers. However it is my opinion that range of absolute error (RAE) should be the reported figure of merit. The total prediction accuracy in future in the field will be the worst possible prediction error and this is the RAE. MAE may be very small but the standard deviation may be large resulting in a trained model that is not very useful. Interestingly the light grey predictions are approximately one human step in duration and this figure shows that one step is not a strong predictor of gait or fatigue in rough terrain. This could be viewed as a latent space with a correct mean but large standard deviation. An accurate prediction is attained by running a low pass filter across the predictions with the effect of narrowing the probability distribution function and giving a reduced range to any one prediction, making the trained model useful in the field. This result is shown for 200 predictions

(102 s) with activity ‘walk up’. It is worth noting the data in Figure 5-5 is not continuous as it is only for activity ‘walk up’ which occurred regularly over the protocol but was broken by different slopes and activities such as run down and climb gate. The protocol did not give accurate results for all activities indicating some activities such as walking on a flat surface is not modulated significantly by fatigue or there was insufficient data for this activity generated in the protocol.

This research used a novel, to the author’s knowledge, approach to determine fatigue from a single sensor (IMU + ECG) by first performing a HAR function to determine activity type, subjecting the participant to an hourly protocol of physical and cognitive load until cessation with hourly assessments. The fatigue ML model used the sensor data plus HAR model output as a preprocessed input to train against the performance assessments and determine which activity was best to use to determine fatigue. Additionally, the prediction latent space was narrowed with a rolling filter to increase the accuracy of any future predictions. To compare 200 predictions with a non-overlapping window of 128 samples is the equivalent of 25,600 neurons in the input layer of a neural network which would require a supercomputer to train and predict.

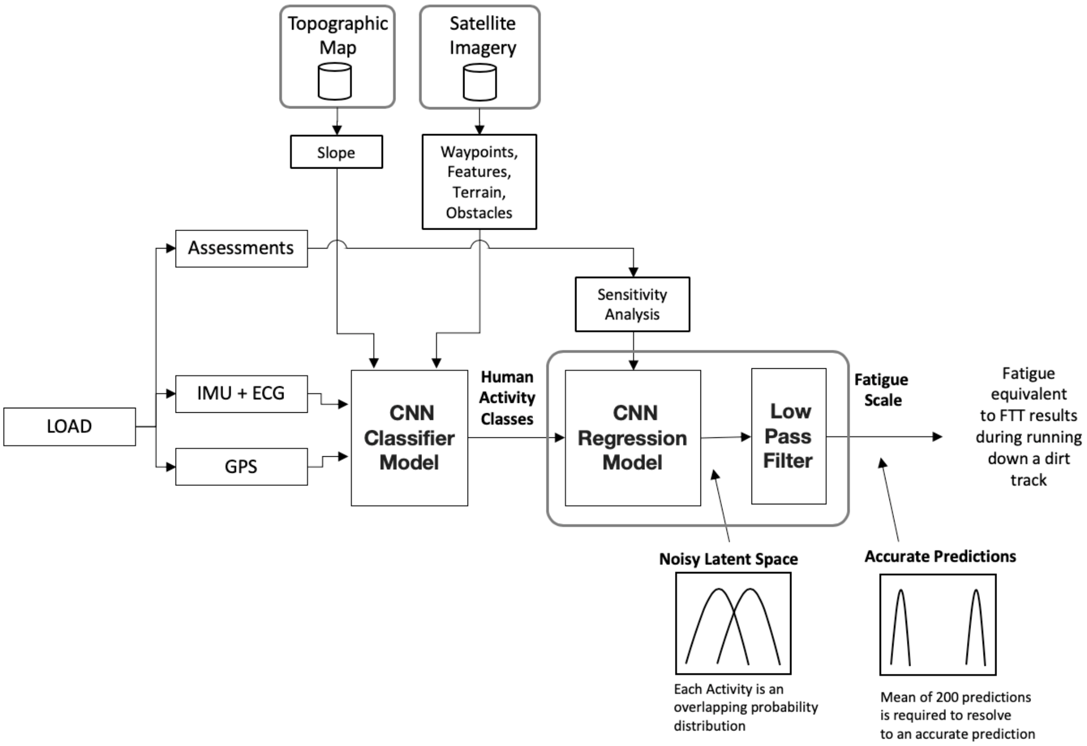


Figure 6-1 A cascaded deep learning model with classification of human activity recognition and fatigue prediction with post prediction to filtering

Protocol

A significant contribution of this work was the protocol which enabled fatigue research to be performed in the field. Several challenges were solved including using satellite imagery and topographical information to remotely calibrate a field environment for slope surface texture and obstacles. The protocol took the field and laboratory assessments, performed over an hourly protocol to enable accurate measurement of fatigue degradation over time. This protocol is useful for further research where different loads can be carried out and also larger numbers of subjects can be used in a safe and monitored way as those subjects will be fatiguing to failure.

Applications

This research and findings has application to sports and exercise science, industrial and military activities in the field where further insight is required into short term and long term fatigue, recovery cycles, which activities lead to these behaviours and what strategies a participant undertakes in the field with different levels of fatigue and goals. A single sensor for these applications is required as the participant is usually loaded with equipment and working near their limit of endurance. Any sensor that adds weight or interferes with operational performance would not be tolerated. These applications require a participant to ambulate over rough terrain while making strategic decisions making gait an appropriate measure. Any sensor in these applications will require sufficient memory and battery power for several days and be mechanically robust and self-contained with no wires or other entanglement risks.

Machine learning models have enabled measurement of cognitive and physical fatigue with a single sensor in a highly variable environment. Without the ML model it is likely more sensors would be required which is not practical in many long duration remote environments. Put another way, adding the sophistication of a deep learning model has reduced the carriage load, by reducing power, memory, size and weight for the individual.

Limitations and Future Work

The findings from the first experiment in this thesis was that the protocol, logistics and data analysis were challenging in remote locations. Repetitive data entry such as name and date of birth lead to low compliance. Daily protocols were insufficient to pick up hourly changes in a person's fatigue or train a ML model. Based on this finding the protocol for the second field trial was refined. Notable decisions included limiting participant numbers and the number and type of fatigue measurements while the protocol was validated. The data analysis took six months of data wrangling. While the dataset had only one participant it did have three times the data points of the WISDM dataset and uniquely included terrain variation and fatigue. Further work is now required with more participants. This work can build on research to date by using activity recognition to speed up data analysis. Additionally further work can be done for longer time periods, with unplanned field routes using the satellite imagery protocol developed. The HAR will need to be validated against more participants which may take additional resources and software development to enable a faster work flow. Planning is required to avoid injuring participants and research into critical power with high cognitive loads to determine personal loading versus duration. Sleep deprivation with longer period trials is required.

In retrospect I would develop a pre-protocol assessment to determine the critical power of a participant and assess their reduction in critical power with and without cognitive load. This would enable an estimation of physical load a participant could take and hence enable them to last 24 hours and generate sleep deprivation.

Secondly considering a track to simplify labelling is worth considering, however the track complexity did lead to insights that make the protocol easier to extend into more remote filed studies. The multidisciplined nature of this research and extending the experiment into the field has taken time and taught many small lessons in process and protocol management.

The research showed fatigue is a complex set of overlapping and redundant systems that include physiology, biomechanics, neurology, psychology. The limiting factor to performance will vary from case to case and person to person. However the levels of fatigue can be measured which may give more insight into understanding fatigue in the field.

The work demonstrated short duration cognitive assessments are feasible in the field and show the required sensitivity for validation and ML model training. This was shown by calculating the R^2 of the trend line across the protocol period.

Protocols in remote environments should be designed for: reduced data processing, labelling, validation, AI modelling (lots of labels). The first field trial had daily assessments and this was insufficient to train a model as the variation was large indicating there was probably hourly variations not measured. The protocol did not have a sustained load which added further uncertainty. The second field trail had hourly assessments and defined regular load based on a repeating hourly protocol. This allowed more insight into small variations including recovery such as before and cognitive assessments. The various assessments were sensitive to different degrees of physical and cognitive loads and this showed as they were performed after the physical load and after the cognitive load during sitting.

Critical power should be adjusted for cognitive loads in the protocol, reducing the physical load in order to reach a 24 hours test with sleep deprivation. The participants had completed the planned 24 hour protocol based on previous training. However when cognitive loads were added, for the first time, into the physical load the perception of fatigue increased significantly. In hind sight this has been generally reported but not for critical power as most races do not have a cognitive component.

Range of absolute error should be the reported figure of merit. Not MAE. The MAE is a good error term for back propagation training of the model, however individual predictions may be large. The advantage of RAE is that it reflects the possible prediction error for any single prediction, which is the intended use of this research.

Low pass filtering can be used to reduce inter prediction error on a trained model. This is one of the most exciting findings of this research and shows how applying ML to physiology research can give insight into the actual physiology not just the data science. This work demonstrated through optimisation that 100 seconds is required in a rough terrain to use gait to determine fatigue. That is, an individual step is not indicative a particular fatigue level.

This research has taught that some activities are modulated by fatigue and some are not. Walk down was the most accurate. This is useful in sports and exercise science to understand strategies and reduce injury. This is an instance of where having a remote analysis capability will assist in sports and exercise science in the field.

Deep learning provides good results and does not require feature extraction for both continuous activity and single obstacle climbing. This is useful as it removes the need to determine what statistical features are required to make fatigue measurements and may be using previously uncomputed features and novel combination of features that will not be possible with traditional analysis. CNN works well with time series data in the presence of large amounts of environmental noise.

It is possible to perform unobserved remote field studies with equivalent accuracy and validation methods to controlled laboratory protocols when an appropriate protocol is followed using various remote datasets.

The insights gathered from this work have provided a protocol and approach to push the boundaries into performance and fatigue research in remote multi day environments. The research determined that gait and vital signs while walking up a slope and running down a slope were indicative of cognitive and physical fatigue. This research has shown AI and a single sensor can accurately predict HAR and fatigue in a mountain environment. However a carefully planned protocol is still required. Future work to automate this could enable scaling for many more users.

7. CONCLUSION

This research has addressed the knowledge gaps in using a single sensor and artificial intelligence for assessing human activity recognition and fatigue in the field with variable terrain, obstacles and a self-pacing participant with voluntary activity types and no observer.

This research asked the question could fatigue be measured in the field. This lead to several related questions: (i) what is the definition of fatigue which is relevant to field activities; (ii) how can fatigue be measured in the field with assessments and sensors (iii) what protocol is practical in the field (iv) how can data from unobserved experiments be used to label data for deep learning and finally (v) can datasets with environmental noise be used with deep learning models to achieve accurate prediction results?

This work has built on the current literature and critically assessed what is required to enter the field while maintaining a link to validated protocols. This work has taken previously reported work in human activity recognition and used deep learning to show datasets with significant environmental modulation can still give accurate results.

The thesis started with a field study on multiple people and showed that protocol development was imperative before scaling to other participants. This lead to new research into a compliance formula to enable understanding and planning of field protocols. The result of the first field study was a redesigned second protocol that was logistically more manageable, enabled validation to gold standard laboratory assessments and a repeated periodic uniform load to enable machine learning training. This work went past traditional machine learning and used deep learning. Where the majority of machine learning research was performed on existing datasets, this work generated a new dataset with unique features and environmental considerations.

The first field study was with multiple participants and showed that a new protocol had many aspects which required testing and refining before scaling to other participants. Additionally the calibration of the field environment and development of data analysis and labelling tools for deep learning takes a large amount of resources. Given these lessons it was determined that the entire approach should be developed and proven first on a single participant as failure would lead to protocol refinement before scaling to more people. The success of the protocol and the accuracy results attained should enable successful funding for future research with additional resources.

A single sensor was imposed as a limiting factor for this research to enable multiple day remote field research, due to size, volume and power constraints. The single sensor approach continues to be a challenge where there may be less data available for the model. However, the results showed accurate results could be obtained when the sensor was combined with an AI model.

A new compliance formula was derived based on the challenges from the first field experiment that can potentially aid further research and protocol design. This compliance formula enables the deconstruction of work effort versus reward for tasks and enables prediction of compliance.

A new protocol was developed to calibrate a mountain trail using satellite images and topographic information that enabled fatigue research in the field. Remotely collected data were labelled to enable a deep learning model to be trained.

It was shown that human activity recognition is possible with a single sensor using an AI classifier model in an uncontrolled environment, variable terrain, obstacles and fatigue up to cessation of exercise. The AI model was optimised for both repetitive and one off activities such as running vs obstacle climbing.

Finally, the overall question of the research was proven that fatigue can be measured in the field using a single sensor with an AI regression model. The model was novel in that is cascaded a HAR classifier and a fatigue regression model followed by a averaging filter to improve prediction accuracy. The research determined that MAE is a good error term for model training, however RAE should be used as a figure of merit.

The research was performed with a unique protocol that enables larger field trials. However, the labour content of data wrangling is not practical and requires development of automation software. Only a semi-automated approach could be used for this research.

Automation of the data wrangling will enable this protocol to be validated with a larger number of people under various terrain types with different fitness levels, motivation levels and different goals. This research may require travelling to an ultra-marathon competition to recruit sufficient participants that have trained for the event.

Fatigue research is important in dangerous and performance oriented activities and protocols are needed that do not distract or require activity cessation. It may be possible to take this work and with further machine learning to research the underlying reasons for fatigue in a given situation. This would allow research of fatigue pathways in the field. This is important work as the literature[4,234,235,245–247] makes it clear that fatigue is a multi-faceted phenomenon encompassing physiology, neurology, psychology, biomechanics, environment and task goals where the factors in the field cannot be replicated in a laboratory environment.

In summary, this thesis showed that fatigue is a complexity of redundant overlapping systems and a definition requires context from the field and mission goal to be applicable. Fatigue should be measured throughout an activity and not be solely focused on end points. Compliance is calculable and needs to be designed into the protocol to be successful in the field. A single sensor when coupled with a deep learning model can accurately classify human activity recognition. Fatigue can be predicted in the field using a single sensor and trained AI model. Further work is required to automate the data pipeline and protocol to enable multiple subject variation to be studied in the field.

The research area of fatigue and performance is increasingly complex as it leaves the laboratory and moves into the field. Once in the field there are various challenges; compliance to protocol and wearing sensors, fatiguing subjects safely, collecting error free sensor data, validation protocols and labelling of data. Many factors may influence fatigue and this requires a multidisciplined approach, as summarized by Nindl[248], when discussing the state of readiness and preparedness for military applications. The work of Noakes[247–249] and Marcora[247] has taken the field to its limits in the laboratory whilst work by Clermont[239] has shown promising results in the field when looking at fatigue induced gait differences.

The use of cognitive assessments in a highly repetitive way for machine learning has been developed during this work. The use of satellite imagery and topography to label and calibrate an outside track for physical load assessment is new and will enable further remote research to be under taken without the need for researchers to be present. Field trials will be possible without prior knowledge of a track taken, such as on a military mission. Further development of the assessment applications will make this work more scalable.

The exciting aspect of this work is that it provides a pathway for further fatigue and performance research. This work has given some insight into short-term and long-term fatigue and recovery. This may enable further research into how to optimize both load and resilience. The ultimate goal is to enable the answer to the question "is this person capable of successfully performing the next mission?" and this work has contributed to the start of a collection of tools that can be used for this goal. This work has led the author into collaboration with researchers and organizations that are investigating humans in harsh environments and that require autonomous tools to manage their wellness and performance.

8. CONFLICTS OF INTEREST

There are no conflicts of interest to declare.

REFERENCES

1. Boyd JR. The Essence of Winning and Losing Key Statements.
2. Von Jan T, Karnahl T, Seifert K, Hilgenstock J, Zobel R. Don't sleep and drive – VW's fatigue detection technology. *Proc 19th Int Conf Enhanc Saf Veh*. 2005:1-12.
3. Reason J. Human error: Models and management. *West J Med*. 2000;172(6):393-396. doi:10.1136/ewjm.172.6.393
4. Boghini G, Astolfi L, Vecchiato G, Mattia D, Babilioni F. Mental fatigue impairs physical performance in humans. *Neurosci Biobehav Rev*. 2014;44(1):58-75. doi:10.1109/CIDM.2009.4938635
5. Shafizadeh M, Crowther R, Ali A, Davids K. Effects of dual task constraints on intra-limb coordination during treadmill walking in people with chronic stroke. 2017;71(2):8-19.
6. Lemke NC, Wiloth S, Werner C, Hauer K. Validity, test-retest reliability, sensitivity to change and feasibility of motor-cognitive dual task assessments in patients with dementia. *Arch Gerontol Geriatr*. 2017;70:169-179. doi:10.1016/j.archger.2017.01.016
7. Kizony R, Levin M, Hughey L, Perez C, Fung J, Perry. Cognitive load and dual-task performance during locomotion poststroke: a feasibility study using a functional virtual environment. *Phys Ther*. 2010;90(2):252-260. http://www.ncbi.nlm.nih.gov/pubmed/20023003.
8. Verlinden VJA, Geest JN Van Der, Hofman A, Ikram MA. Cognition and gait show a distinct pattern of association in the general population. *Alzheimer's Dement*. 2014;10(3):328-335. doi:10.1016/j.jalz.2013.03.009
9. Moon HI, Pyun SB, Tae WS, Kwon HK. Neural substrates of lower extremity motor, balance, and gait function after supratentorial stroke using voxel-based lesion symptom mapping. *Neuroradiology*. 2016;58(7):723-731. doi:10.1007/s00234-016-1672-3
10. Theill N, Martin M, Schumacher V, Bridenbaugh SA, Kressig RW. Simultaneously measuring gait and cognitive performance in cognitively healthy and cognitively impaired older adults: The basel motor-cognition dual-task

paradigm. *J Am Geriatr Soc*. 2011;59(6):1012-1018. doi:10.1111/j.1532-5415.2011.03429.x

11. Noakes TD. Challenging beliefs: ex Africa semper aliquid novi. *Med Sci Sports Exerc*. 1997;29(5):571-590. doi:10.1097/00005768-199705000-00001
12. Di Giulio C, Daniele F, Tipton CM. Angelo Mosso and muscular fatigue: 116 Years after the first congress of physiologists: IUPS commemoration. *Am J Physiol - Adv Physiol Educ*. 2006;30(2):51-57. doi:10.1152/advan.00041.2005
13. Millet GY. Can neuromuscular fatigue explain running strategies and performance in ultra-marathons?: The flush model. *Sport Med*. 2011;41(6):489-506. doi:10.2165/11588760-000000000-00000
14. Marcora SM, Staiano W, Manning V. Mental fatigue impairs physical performance in humans. *J Appl Physiol Publ*. 2009;106(3):857-864. doi:10.1152/japplphysiol.91324.2008
15. Hale T. *History of Developments in Sport and Exercise Physiology: A. V. Hill, Maximal Oxygen Uptake, and Oxygen Debt*. Vol 26.; 2008. doi:10.1080/02640410701701016
16. Hill AV, Long CNH, Lupton H. Muscular Exercise Lactic Acid and the Supply and Utilisation of Oxygen. *R Soc*. 1924;97(682):155-176.
17. Wasserman K, Whipp B, Koyal S, Beaver W. Anaerobic threshold and respiratory gas exchagne during exercise. *J Appl Physiol*. 1973;35(2):236-243. https://www.golder.com/insights/block-caving-a-viable-alternative/.
18. Calatayud J, Alberton CL, Colado JC, et al. Noninvasive Determination of Anaerobic Threshold Based on the Heart Rate Deflection Point in Water Cycling. *J Strength Cond Res*. 2015;30(2):518-524. doi:10.1519/jsc.0000000000001099
19. F. Conconi, G. Grazzi, C. Guglielmini, C. Borsetto, E. Ballarin, G Mazzoni, M. Patracchini FM. Conconi 1996 The Conconi Test Methodology after 12 yers of application. 1996.
20. Melzack R, Wall P. Pain mechanisms: a new theory. *Science (80-)*. 1965;150(3699):977-979.
21. Melzack R. Gate control theory. *Pain Forum*. 1996;5(2):128-138. doi:10.1016/s1082-3174(96)80050-x
22. Klinger E. Consequences of commitment to and disengagement from incentives. *Psychol Rev*. 1975;82(1):1-23. doi:10.1093/mind/xxv.2.280

23. Brandstätter V, Schüler J. Action crisis and cost–benefit thinking: A cognitive analysis of a goal-disengagement phase. *J Exp Soc Psychol*. 2013;49(3):543-553. doi:10.1016/j.jesp.2012.10.004

24. Borg G. Psychophysical bases of perceived exertion. *Med Sci Sport Exerc*. 1982;14(5):377 381.

25. Enoka RM, Stuart DG. Neurobiology of muscle fatigue. *J Appl Physiol*. 1992;72(5):1631-1648. doi:10.1152/jappl.1992.72.5.1631

26. St. Clair Gibson A, Noakes TD. Evidence for complex system integration and dynamic neural regulation of skeletal muscle recruitment during exercise in humans. *Br J Sports Med*. 2004;38(6):797-806. doi:10.1136/bjsm.2003.009852

27. Vanhatalo A, Jones AM, Burnley M. Application of Critical Power in Sport What Is the Critical Power Concept. *Int J Sports Physiol Perform*. 2011;6:128-136.

28. Enoka R, Duchateau J. Translating fatigue to human performance. *Meical Sci Sport Exerc*. 2016;48(11):2223-2238. doi:10.1016/j.physbeh.2017.03.040

29. Enoka RM, Baudry S, Rudroff T, Farina D, Klass M, Duchateau J. Unraveling the neurophysiology of muscle fatigue. *J Electromyogr Kinesiol*. 2011;21(2):208-219. doi:10.1016/j.jelekin.2010.10.006

30. Noakes TD. Fatigue is a brain-derived emotion that regulates the exercise behavior to ensure the protection of whole body homeostasis. *Front Physiol*. 2012:82. doi:10.3389/fphys.2012.00082

31. Venhorst A, Micklewright D, Noakes TD. Perceived Fatigability : Utility of a Three-Dimensional Dynamical Systems Framework to Better Understand the Psychophysiological Regulation of Goal-Directed Exercise Behaviour. *Sport Med*. 2018:2479-2495.

32. Wylie GR, Dobryakova E, Deluca J, Chiaravalloti N, Essad K, Genova H. Cognitive fatigue in individuals with traumatic brain injury is associated with caudate activation. *Sci Rep*. 2017;7(1). doi:10.1038/s41598-017-08846-6

33. Boksem MAS, Tops M. Mental fatigue: Costs and benefits. *Brain Res Rev*. 2008;59(1):125-139. doi:10.1016/j.brainresrev.2008.07.001

34. Möckel T, Beste C, Wascher E. The Effects of Time on Task in Response Selection - An ERP Study of Mental Fatigue. *Nat Publ Gr*.:1-9. doi:10.1038/srep10113

35. Borghini G, Astolfi L, Vecchiato G, Mattia D, Babiloni F. Measuring

neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci Biobehav Rev*. 2014;44:58-75. doi:10.1016/j.neubiorev.2012.10.003

36. Aoyagi MW, Portenga ST. The Role of Positive Ethics and Virtues in the Context of Sport and Performance Psychology Service Delivery. *Prof Psychol Res Pract*. 2010;41(3):253-259. doi:10.1037/a0019483
37. Barker JB, Neil R, Fletcher D. Using Sport and Performance Psychology in the Management of Change. *J Chang Manag*. 2016;16(1):1-7. doi:10.1080/14697017.2016.1137149
38. Aidman E. Cognitive Fitness Framework: Towards Assessing, Training and Augmenting Individual-Difference Factors Underpinning High-Performance Cognition. *Front Hum Neurosci*. 2020;13(January):1-9. doi:10.3389/fnhum.2019.00466
39. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proc Hum Factors Ergon Soc Annu Meet*. 2006;50(9):904-908. doi:10.1177/154193120605000909
40. Chen LL, Zhao Y, Zhang J, Zou JZ. Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning. *Expert Syst Appl*. 2015;42(21):7344-7355. doi:10.1016/j.eswa.2015.05.028
41. Powell DMC, Spencer MB, Petrie KJ. Automated collection of fatigue ratings at the top of descent: A practical commercial airline tool. *Aviat Sp Environ Med*. 2011;82(11):1037-1041. doi:10.3357/ASEM.3115.2011
42. Asmaro D, Mayall J, Ferguson S. Cognition at altitude: Impairment in executive and memory processes under hypoxic conditions. *Aviat Sp Environ Med*. 2013;84(11):1159-1165. doi:10.3357/ASEM.3661.2013
43. Caldwell JA, Mallis MM, Caldwell JL, Paul MA, Miller JC, Neri DF. Fatigue countermeasures in aviation. *Aviat Sp Environ Med*. 2009;80(1):29-59. doi:10.3357/ASEM.2435.2009
44. Bendak S, Rashid HSJ. International Journal of Industrial Ergonomics Fatigue in aviation : A systematic review of the literature. *Int J Ind Ergon*. 2020;(76).
45. Dehghani F, Golbabaei F, Omidi F, Zakerian SA. Investigation of the effect of unusual work shifts and sleeps deprivation on cognitive performance in workers in an automotive industry. *Iran Occup Heal*. 2019;16(3):26-35.

46. Shortz AE, Mehta RK, Peres SC, Benden ME, Zheng Q. Development of the fatigue risk assessment and management in high-risk environments (FRAME) survey: A participatory approach. *Int J Environ Res Public Health*. 2019;16(4). doi:10.3390/ijerph16040522

47. Kim S, Ros R, Hussein R. Detecting Fatigue Driving Through PERCLOS : A Review. *Int J Image Process*. 2020;14(1):1-7.

48. Murugan S, Selvaraj J, Sahayadhas A. Analysis of different measures to detect driver states: A review. *2019 IEEE Int Conf Syst Comput Autom Networking, ICSCAN 2019*. 2019;(i). doi:10.1109/ICSCAN.2019.8878844

49. Liu F, Li X, Lv T, Xu F. A Review of Driver Fatigue Detection: Progress and Prospect. *2019 IEEE Int Conf Consum Electron ICCE 2019*. 2019. doi:10.1109/ICCE.2019.8662098

50. Techera U, Hallowell M, Littlejohn R. Worker Fatigue in Electrical-Transmission and Distribution-Line Construction. *J Constr Eng Manag*. 2019;145(1):1-9. doi:10.1061/(ASCE)CO.1943-7862.0001580

51. Noble DD. Cockpit cognition: Education, the military and cognitive engineering. *AI Soc*. 1989;3(4):271-296. doi:10.1007/BF01908619

52. Basner M, Savitt A, Moore TM, et al. Development and Validation of the Cognition Test Battery for Space Flight. *Aerosp Med Hum Perform*. 2016;86(11):942-952. doi:10.3357/AMHP.4343.2015.Development

53. Soltani S, Mutka M. ArgMax and ArgMin transitional probabilistic models in cognitive radio mesh networks. *Wirel Commun Mob Comput*. 2015;15:1355-1367. doi:10.1002/wcm

54. Eggleton P, Eggleton GP. The Inorganic Phosphate and a Labile Form of Organic Phosphate in the Gastrocnemius of the Frog. *Biochem J*. 1927;21(1):190-195. doi:10.1042/bj0210190

55. Langen P, Hucho F. Karl Lohmann and the discovery of ATP. *Angew Chemie - Int Ed*. 2008;47(10):1824-1827. doi:10.1002/anie.200702929

56. Pimenta A, Carneiro D, Neves J, Novais P. A neural network to classify fatigue from human-computer interaction. *Neurocomputing*. 2016;172:413-426. doi:10.1016/j.neucom.2015.03.105

57. Abdulin E. User Fatigue Detection via Eye Movement Behavior. 2015.

58. Gonzalez K, Sasangohar F, Mehta R, Lawley M EM. Measuring Fatigue through Heart Rate Variability and Activity Recognition: A Scoping Literature Review

of Machine Learning Techniques. *Proc Hum Factors Ergon Soc Annu Meet*. 2017:1748.

59. Patel AN, Howard MD, Roach SM, et al. Mental State Assessment and Validation Using Personalized Physiological Biometrics. *Front Hum Neurosci*. 2018;12(June):221. doi:10.3389/fnhum.2018.00221

60. Azim T, Jaffar MA, Mirza AM. Fully automated real time fatigue detection of drivers through Fuzzy Expert Systems. *Appl Soft Comput J*. 2014;18:25-38. doi:10.1016/j.asoc.2014.01.020

61. Aryal A, Ghahramani A, Becerik-Gerber B. Monitoring fatigue in construction workers using physiological measurements. *Autom Constr*. 2017;82:154-165. doi:10.1016/j.autcon.2017.03.003

62. Zhang J, Yu B, Nguyen LT, et al. Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors. *Int Conf Mob Comput Appl Serv*. 2014. doi:10.4108/icst.mobicase.2014.257786

63. Kirchner M, Schubert P, Liebherr M, Haas CT. Detrended fluctuation analysis and adaptive fractal analysis of stride time data in Parkinson’s disease: Stitching together short gait trials. *PLoS One*. 2014;9(1):1-6. doi:10.1371/journal.pone.0085787

64. Schurgin MW, Nelson J, Iida S, Ohira H, Chiao JY, Franconeri SL. Eye movements during emotion recognition in faces. *J Vis*. 2014;14(13):14-14. doi:10.1167/14.13.14

65. Dinges DF, Mallis MM, Maislin G, Powell IV JW. Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and as the Basis for Alertness Management. 1998;(April):113. doi:10.1037/e496472008-001

66. Guidi A, Greco A, Felici F, et al. Heart rate variability analysis during muscle fatigue due to prolonged isometric contraction. *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS*. 2017:1324-1327. doi:10.1109/EMBC.2017.8037076

67. Brindle RC, Ginty AT, Phillips AC, Fisher JP, Mcintyre D, Carroll D. Heart rate complexity: A novel approach to assessing cardiac stress reactivity. *Psychophysiology*. 2016;53(4):465-472. doi:10.1111/psyp.12576

68. Grobe S, Kakar RS, Smith ML, Mehta R, Baghurst T, Boolani A. Impact of cognitive fatigue on gait and sway among older adults: A literature review. *Prev Med Reports*. 2017;6:88-93. doi:10.1016/j.pmedr.2017.02.016

69. Heredia-Jimenez J, Latorre-Roman P, Santos-Campos M, Orantes-Gonzalez E,

Soto-Hermoso VM. Spatio-temporal gait disorder and gait fatigue index in a six-minute walk test in women with fibromyalgia. *Clin Biomech*. 2016;33:1-6. doi:10.1016/j.clinbiomech.2016.01.009

70. Fullagar HHK, Skorski S, Duffield R, Hammes D, Coutts AJ, Meyer T. Sleep and Athletic Performance: The Effects of Sleep Loss on Exercise Performance, and Physiological and Cognitive Responses to Exercise. *Sport Med*. 2015;45(2):161-186. doi:10.1007/s40279-014-0260-0

71. Van Dongen HPA, Dinges DF. Sleep, circadian rhythms, and psychomotor vigilance. *Clin Sports Med*. 2005;24(2):237-249. doi:10.1016/j.csm.2004.12.007

72. Fimm B, Brand T, Spijkers W. Time-of-day variation of visuo-spatial attention. *Br J Psychol*. 2016;107(2):299-321. doi:10.1111/bjop.12143

73. Granacher U, Wolf I, Wehrle A, Bridenbaugh S, Kressig RW. Effects of muscle fatigue on gait characteristics under single and dual-task conditions in young and older adults. *J Neuroeng Rehabil*. 2010;7(1):1-12. doi:10.1186/1743-0003-7-56

74. Fuller JT, Bellenger CR, Thewlis D, et al. Tracking performance changes with running-stride variability when athletes are functionally overreached. *Int J Sports Physiol Perform*. 2017;12(3):357-363. doi:10.1123/ijspp.2015-0618

75. Attal F, Mohammed S, Dedabrishvili M, Chamroukhi F, Oukhellou L, Amirat Y. Physical Human Activity Recognition Using Wearable Sensors. *Sensors*. 2015;15(12):31314-31338. doi:10.3390/s151229858

76. Simoneau M, Bégin F, Teasdale N. The effects of moderate fatigue on dynamic balance control and attentional demands. *J Neuroeng Rehabil*. 2006;3:22. doi:10.1186/1743-0003-3-22

77. Hsu B-W, Wang M-JJ, Chen C-Y, Chen F. Effective Indices for Monitoring Mental Workload While Performing Multiple Tasks. *Percept Mot Skills*. 2015;121(1):94-117. doi:10.2466/22.PMS.121c12x5

78. Wascher E, Rasch B, S??nger J, et al. Frontal theta activity reflects distinct aspects of mental fatigue. *Biol Psychol*. 2014;96(1):57-65. doi:10.1016/j.biopsycho.2013.11.010

79. Barwick F, Arnett P, Slobounov S. EEG correlates of fatigue during administration of a neuropsychological test battery. *Clin Neurophysiol*. 2012;123(2):278-284. doi:10.1016/j.clinph.2011.06.027

80. Eckstein MK, Guerra-carrillo B, Miller AT, Bunge SA. Developmental Cognitive Neuroscience Beyond eye gaze : What else can eyetracking reveal

about cognition and cognitive development ? *Dev Cogn Neurosci*. 2017;25:69-91.

81. Hooge I, Camps G. Scan path entropy and arrow plots: Capturing scanning behavior of multiple observers. *Front Psychol*. 2013;4(DEC):1-10. doi:10.3389/fpsyg.2013.00996
82. Montero-Odasso M, Muir SW, Speechley M. Dual-task complexity affects gait in people with mild cognitive impairment: The interplay between gait variability, dual tasking, and risk of falls. *Arch Phys Med Rehabil*. 2012;93(2):293-299. doi:10.1016/j.apmr.2011.08.026
83. Chua EC-P, Tan W-Q, Yeo S-C, et al. Heart Rate Variability Can Be Used to Estimate Sleepiness-related Decrements in Psychomotor Vigilance during Total Sleep Deprivation. *Sleep*. 2012. doi:10.5665/sleep.1688
84. Luque-Casado A, Perales JC, Cárdenas D, Sanabria D. Heart rate variability and cognitive processing: The autonomic response to task demands. *Biol Psychol*. 2015;113:83-90. doi:10.1016/j.biopsycho.2015.11.013
85. Mizuno K, Tanaka M, Yamaguti K, Kajimoto O, Kuratsune H, Watanabe Y. Mental fatigue caused by prolonged cognitive load associated with sympathetic hyperactivity. *Behav Brain Funct*. 2011;7(1):17. doi:10.1186/1744-9081-7-17
86. Roelands B, De Koning J, Foster C, Hettinga F, Meeusen R. Neurophysiological determinants of theoretical concepts and mechanisms involved in pacing. *Sport Med*. 2013;43(5):301-311. doi:10.1007/s40279-013-0030-4
87. Iodice P, Calluso C, Barca L, Bertollo M, Ripari P, Pezzulo G. Fatigue increases the perception of future effort during decision making. *Psychol Sport Exerc*. 2017;33:150-160. doi:10.1016/j.psychsport.2017.08.013
88. McLaughlin AC, Simon DA, Gillan DJ. From Intention to Input: Motor Cognition, Motor Performance, and the Control of Technology. *Rev Hum Factors Ergon*. 2010;6(1):123-171. doi:10.1518/155723410x12849346788741
89. James W (Harvard U. *James, 1890 - The Principles of Psychology.Pdf.*; 1890.
90. Atkinson RC, Shiffrin RM. Human memory a proposed system and its control processes. *Psychol Learn Motiv*. 1968:89-195.
91. Baddeley AD, Hitch G. Working memory. *Appl Psychol*. 1974:47-89.
92. Ericsson KA. Towards a science of the acquisition of expert performance in sports: Clarifying the differences between deliberate practice and other types of practice. *J Sports Sci*. 2020;38(2):159-176.

doi:10.1080/02640414.2019.1688618

93. Bier L, Wolf P, Hilsenbek H, Abendroth B. How to measure monotony-related fatigue. *Theor Issues Ergon Sci*. 2020;21:22-55.

94. Arrieux J, Cole W, Ahrens A. A review of the validity of computerized neurocognitive assessment tools in mild traumatic brain injury assessment.pdf. *Futur Med*. 2017;2(1).

95. Janssen J, Koekkoek PS, Moll Van Charante EP, Jaap Kappelle L, Biessels GJ, Rutten GEHM. How to choose the most appropriate cognitive test to evaluate cognitive complaints in primary care. *BMC Fam Pract*. 2017;18(1):1-8. doi:10.1186/s12875-017-0675-4

96. Basner M, Savitt A, Moore TM, et al. Development and Validation of the Cognition Test Batery for Spaceflight. *Aerosp Med Hum Perform*. 2015;86(11):942-952. doi:10.3357/AMHP.4343.2015.Development

97. Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery , CNS Vital Signs. *Arch Clin Neuropsychol*. 2006;21:623-643. doi:10.1016/j.acn.2006.05.007

98. Memorandum NT, Comstock JR, Arnegard RJ. The Multi Attribute Task Battery for Human Operator Workload and Strategic Behaviour Research. *Nasa Tech Memo*. 1992;(January 1992).

99. Cherry BJ, Zettel-Watson L, Chang JC, Shimizu R, Rutledge DN, Jones CJ. Positive associations between physical and cognitive performance measures in fibromyalgia. *Arch Phys Med Rehabil*. 2012;93(1):62-71. doi:10.1016/j.apmr.2011.08.006

100. Lee CY, Kang SJ, Hong SK, Ma H Il, Lee U, Kim YJ. A validation study of a smartphone-based finger tapping application for quantitative assessment of bradykinesia in Parkinson’s disease. *PLoS One*. 2016;11(7):1-11. doi:10.1371/journal.pone.0158852

101. Leyla A, Kiziltan E. Polyphasic Temporal Behavior of Finger-Tapping Performance : *J Mot Behav*. 2016;48(1):72-78.

102. Pageaux B, Marcora SM, Rozand V, Lepers R. Mental fatigue induced by prolonged self-regulation does not exacerbate central fatigue during subsequent whole-body endurance exercise. *Front Hum Neurosci*. 2015;9(February):1-12. doi:10.3389/fnhum.2015.00067

103. Ferreira IS, Simões MR, Marôco J. The Addenbrooke’s Cognitive Examination

Revised as a potential screening test for elderly drivers. *Accid Anal Prev*. 2012;49:278-286. doi:10.1016/j.aap.2012.03.036

104. Kluckow SW, Rehbein JG, Schwab M, Witte OW, Bublak P. What you get from what you see: Parametric assessment of visual processing capacity in multiple sclerosis and its relation to cognitive fatigue. *Cortex*. 2016;83:167-180. doi:10.1016/j.cortex.2016.07.018
105. Godi M, Franchignoni F, Caligari M, Giordano A, Turcato AM, Nardone A. Comparison of Reliability, Validity, and Responsiveness of the Mini- BESTest and Berg Balance Scale in Patients With Balance Disorders. *Phys Ther*. 2018;93(2):158.
106. Theill N, Martin M, Schumacher V, Bridenbaugh SA, Kressig RW. Simultaneously measuring gait and cognitive performance in cognitively healthy and cognitively impaired older adults: The basel motor-cognition dual-task paradigm. *J Am Geriatr Soc*. 2011;59(6):1012-1018. doi:10.1111/j.1532-5415.2011.03429.x
107. Freedman M, Leach L, Kaplan E, Winocor G, Shutman K, Delis D. Clock drawing An neuropsychological analysis. *J Pers Assess*. 1996:439.
108. Nyborn JA, Himali JJ, Beiser AS, et al. The Framingham heart Study Clock Drawing Performance Normative Data from the offspring cohort. 2013;39(1):617-638. doi:10.1080/0361073X.2013.741996.The
109. Homack S, Riccio CA. A meta-analysis of the sensitivity and specificity of the Stroop Color and Word Test with children. *Arch Clin Neuropsychol*. 2004;19(6):725-743. doi:10.1016/j.acn.2003.09.003
110. van der Linden D, Frese M, Meijman TF. Mental fatigue and the control of cognitive processes: Effects on perseveration and planning. *Acta Psychol (Amst)*. 2003;113(1):45-65. doi:10.1016/S0001-6918(02)00150-6
111. Wu J, Han Y, Zhao FL, Zhou J, Chen Z, Sun H. Validation and comparison of EuroQoL-5 dimension (EQ-5D) and Short Form-6 dimension (SF-6D) among stable angina patients. *Health Qual Life Outcomes*. 2014;12(1):1-11. doi:10.1186/s12955-014-0156-6
112. Amer M, Hubert G, Sullivan SJ, Herbison P, Franz EA, Hammond-Tooke GD. Reliability and diagnostic characteristics of clinical tests of upper limb motor function. *J Clin Neurosci*. 2012;19(9):1246-1251. doi:10.1016/j.jocn.2011.12.007

113. Lucas SJE, Anson JG, Palmer CD, Hellemans IJ, Cotter JD. The impact of 100 hours of exercise and sleep deprivation on cognitive function and physical capacities. *J Sports Sci*. 2009;27(7):719-728. doi:10.1080/02640410902798167

114. Kong W, Zhou Z, Jiang B, Babiloni F, Borghini G. Assessment of driving fatigue based on intra/inter-region phase synchronization. *Neurocomputing*. 2017;219(August 2016):474-482. doi:10.1016/j.neucom.2016.09.057

115. Kaida K, Takahashi M, ??kerstedt T, et al. Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clin Neurophysiol*. 2006;117(7):1574-1581. doi:10.1016/j.clinph.2006.03.011

116. Van Der Elst W, Van Boxtel M, Van Breukelen G, Jolles J. The Letter Digit Substitution Test: Normative data for 1,858 healthy participants aged 24-81 from the Maastricht Aging Study (MAAS): Influence of age, education, and sex. *J Clin Exp Neuropsychol*. 2006;28(6):998-1009. doi:10.1080/13803390591004428

117. Orasanu J, Parke B, Kraft N, et al. Evaluating the effectiveness of schedule changes for Air Traffice Service (ATS) Providers: Controller alertness and fatigue monitoring study (No. DOT/FAA/HFD-13/001). 2012.

118. Vera-villarroel P, Urzu A, Jaime D, et al. Positive and Negative Affect Schedule (PANAS): Psychometric Properties and Discriminative Capacity in Several Chilean Samples. 2019;42(4):473-497. doi:10.1177/0163278717745344

119. Tombaugh TN. A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). 2006;21:53-76. doi:10.1016/j.acn.2005.07.006

120. Fos LA, Greve KW, South MB, et al. Paced Visual Serial Addition Test: An Alternative Measure of Information Processing Speed. *Branchline*. 1994;4282. doi:10.1207/S15324826AN0703

121. Buysse DJ, Reynolds CF, Monk TH, et al. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res*. 1989;28(2):193-213. doi:10.1016/0165-1781(89)90047-4

122. Kos D, Kerckhofs E, Carrea I, Verza R, Ramos M, Jansa J. Evaluation of the Modified Fatigue Impact Scale in four different European countries. *Mult Scler J*. 2005;11(1):76-80. doi:10.1191/1352458505ms1117oa

123. Memorandum NT, Comstock JR, Arnegard RJ. The multi attribute task battery for human operator workload and strategic behaviour research. 1992;(January 1992).

124. Tombaugh TN, McIntyre NJ. The Mini-Mental State Examination: A Comprehensive Review. *J Am Geriatr Soc*. 1992;40(9):922-935. doi:10.1111/j.1532-5415.1992.tb01992.x

125. Carskadon M a, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. *Sleep*. 1986;9(4):519-524. doi:10.1210/JC.2003-031562

126. Kirchner WK. Age differences in short tem retention of rapidly chaning information. *Psychology*. 1958;SS(4):352-357.

127. Sartang GA. Evaluation of Rating Scale Mental Effort (RSME) effectiveness for mental workload assessment in nurses. *Orig Artic Autumn*. 2016;211(54):211-217.

128. Thomas LC, Gast C, Grube R, Craig K. Fatigue Detection in Commercial Flight Operations: Results Using Physiological Measures. *Procedia Manuf*. 2015;3(Ahfe):2357-2364. doi:10.1016/j.promfg.2015.07.383

129. Conway ARA, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW. Working memory span tasks: A methodological review and user's guide. *Psychon Bull Rev*. 2005;12(5):769-786. doi:10.3758/BF03196772

130. Egner T, Hirsch J. The neural correlates and functional integration of cognitive control in a Stroop task. *Neuroimage*. 2005;24(2):539-547. doi:10.1016/j.neuroimage.2004.09.007

131. Jensen AR. Scoring the Stroop Test 1. *Acta Psyc*. 1965;10(24):398-408.

132. Gonzales JU, James CR, Yang HS, et al. Different cognitive functions discriminate gait performance in younger and older women: A pilot study. *Gait Posture*. 2016;50:89-95. doi:10.1016/j.gaitpost.2016.08.021

133. Jain A, Bansal R, Kumar A, Singh K. A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *Int J Appl Basic Med Res*. 2015;5(2):124. doi:10.4103/2229-516X.157168

134. Bajaj JS, Heuman DM, Sterling RK, et al. Validation of EncephalApp, Smartphone-Based Stroop Test, for the Diagnosis of Covert Hepatic Encephalopathy. *Clin Gastroenterol Hepatol*. 2015;13(10):1828-1835. doi:10.1016/j.cgh.2014.05.011

135. Iancheva D, Trenova AG, Terziyski K, Kandilarova S, Mantarova S. Translational validity of PASAT and the effect of fatigue and mood in patients

with relapsing remitting MS: A functional MRI study. *J Eval Clin Pract*. 2018;24(4):832-838. doi:10.1111/jep.12913

136. Whaley M. Perceived Exertion. *Med &amp Sci Sport &amp Exerc*. 2003;29(3):425. doi:10.1097/00005768-199703000-00022
137. Linthorne NP. Analysis of standing vertical jumps using a force platform. *Am J Phys*. 2001;69(11):1198-1204. doi:10.1119/1.1397460
138. Watkins CM, Barillas SR, Wong MA, et al. Determination of vertical jump as a measure of neuromuscular readiness and fatigue. *J Strength Cond Res*. 2017;31(12):3305-3310. doi:10.1519/JSC.0000000000002231
139. Stroop JR. Studies of Interference in Serial Verbal Reactions. *J Exp Psychol*. 1935;XVIII(6).
140. Muller G, Shumann F. Experimentelle Beiträge zur Untersuchung. *Z Psychol*. 1894:96.
141. Comalli P, Wapner S, Werner H. Interference Effects of Stroop Color-Word Test in Childhood, Adulthood, and Aging. *J Genet Psychol*. 1962;93.
142. Lang CE, Wagner JM, Dromerick AW, Edwards DF. Measurement of Upper-Extremity Function Early After Stroke: Properties of the Action Research Arm Test. *Arch Phys Med Rehabil*. 2006;87(12):1605-1610. doi:10.1016/j.apmr.2006.09.003
143. Aydin L, Kiziltan E, Gundogan N. Polyphasic Temporal Behavior of Finger-Tapping Performance. *J Mot Behav*. 2016;48(1).
144. Bossers WJR, van der Woude LH V, Boersma F, Scherder EJA, van Heuvelen MJG. Recommended measures for the assessment of cognitive and physical performance in older patients with dementia: a systematic review. *Dement Geriatr Cogn Dis Extra*. 2012;2(1):589-609. doi:10.1159/000345038
145. Wilkinson R. Some Factors influencing the effect of environmental stressors upon performance. *Psychol Bull*. 1969;72(4):260-272.
146. Jóhannsdóttir KR, Magnúsdóttir EH, Sigurjónsdóttir S, Guðnason J. The role of working memory capacity in cardiovascular monitoring of cognitive workload. *Biol Psychol*. 2018;132(December 2017):154-163. doi:10.1016/j.biopsycho.2017.12.001
147. Corsi PM. Memory and the medial temporal region of the brain. *PhD Thesis, Dep Psychol McGill Univ*. 1972.
148. Brunetti R, Gatto C Del, Delogu F. eCorsi : implementation and testing of the

Corsi block-tapping task for digital tablets. *Front Psychol*. 2014;5(September):1-8. doi:10.3389/fpsyg.2014.00939

149. Claessen MHG, Ham IJM Van Der. Computerization of the Standard Corsi Block-Tapping Task Affects Its Underlying Cognitive Concepts : A Pilot Study. *Appl Neuropsychol*. 2015;22:180-188. doi:10.1080/23279095.2014.892488
150. Skurvydas A, Kamandulis S, Stanislovaitis A, Mamkus G, Mickevičiené D. Effect of four jumping endurance trainings on metabolic fatigue and on indirect symptoms of skeletal muscle damage. *Biol Sport*. 2010;27(4):255-261.
151. Abdelkarim O, Ammar A, Chtourou H, et al. Relationship between motor and cognitive learning abilities among primary school-aged children. *Alexandria J Med*. 2017;(January):1-7. doi:10.1016/j.ajme.2016.12.004
152. Clarke N, Farthing JP, Lanovaz JL, Krentz JR. Direct and indirect measurement of neuromuscular fatigue in Canadian football players. *Appl Physiol Nutr Metab*. 2015;40(5):464-473. doi:10.1139/apnm-2014-0465
153. Marrier B, Meur Y Le, Robineau J, et al. Quantifying neuromuscular fatigue induced by an intense training session in rugby sevens. *Int J Sports Physiol Perform*. 2017;12(2):218-223. doi:10.1123/ijspp.2016-0030
154. Angus RG, Heslegrave RJ. Effects of sleep loss on sustained cognitive performance during a command and control simulation. *Behav Res Methods Instruments Comput*. 1985;17(1):55-67. doi:10.3758/BF03200897
155. Miller E. A standardised laboratory means of determining suseptibility to coriolis sickness. 1969;35:895-897. doi:10.1038/NNANO.2009.177
156. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro CM, Colantonio A. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: A systematic review and meta-analysis. *Sleep Med Rev*. 2016;25:52-73. doi:10.1016/j.smrv.2015.01.009
157. Rani M, Arumugam G. An Efficient Gait Recognition System For Human Identification Using Modified ICA. *Int J Comput Sci …*. 2010;2:55-67. http://www.airccse.org/journal/jcsit/0210ijcsit4.pdf.
158. Zhang K, Werner P, Sun M, Xavier Pi-Sunyer F, Boozer CN. Measurement of Human Daily Physical Activity. 2003. doi:10.1038/oby.2003.7
159. Ahamed N, Kobsar D, Benson L, et al. Using wearable sensors to classify subject specific biomechanical gait patterns based on changes in environmental weather conditions. *PLoS One*. 2018.

160. Westhuizen J van der, Lasenby J. A review of machine learning applied to time series. 2016;(September).

161. Ignatov A. Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Appl Soft Comput J*. 2018;62:915-922. doi:10.1016/j.asoc.2017.09.027

162. Bai S, Kolter JZ, Koltun V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. 2018. http://arxiv.org/abs/1803.01271.

163. Narayanan A, Desai F, Stewart T, Duncan S, Mackay L. Application of Raw Accelerometer Data and Machine-Learning Techniques to Characterize Human Movement Behavior : A Systematic Scoping Review. 2020.

164. Cust EE, Sweeting AJ, Ball K, et al. Machine and deep learning for sport-specific movement recognition : a systematic review of model development and performance. 2019;0414. doi:10.1080/02640414.2018.1521769

165. Beeck T, Meert W, Schutte K, Vanwanseele B, Davis J. Fatigue Prediction in Outdoor Runners Via Machine Learning and Sensor Fusion.pdf. *Appl Data Sci*. 2018:606-615.

166. Lonini L, Gupta A, Kording K, Jayaraman A. Activity recognition in patients with lower limb impairments: Do we need training data from each patient? *2016 38th Annu Int Conf IEEE Eng Med Biol Soc*. 2016:3265.

167. Tripathi S, Acharya S, Sharma RD, Mittal S, Bhattacharya S. Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset. *Twenty-Ninth IAAI Conf*. 2017:4746-4752. https://www.aaai.org/ocs/index.php/IAAI/IAAI17/paper/viewPaper/15007.

168. Laurent F, Valderrama M, Besserve M, et al. Multimodal information improves the rapid detection of mental fatigue. *Biomed Signal Process Control*. 2013;8(4):400-408. doi:10.1016/j.bspc.2013.01.007

169. Nweke HF, Wah TY, Mujtaba G, Algaradi MA. Data fusion and multiple classifier systems for human activity detection and health monitoring: review and open reserach directions. *Inf Fusion*. 2018;(46):147-170.

170. Narayanan A, Desai F, Stewart T, Duncan S, MacKay L. Application of raw accelerometer data and machine-learning techniques to characterize human movement behavior: A systematic scoping review. *J Phys Act Heal*. 2020;17(3):360-383. doi:10.1123/jpah.2019-0088

171. Buckley C, O'Reilly MA, Whelan D, et al. Binary classification of running fatigue using a single inertial measurement unit. *2017 IEEE 14th Int Conf Wearable Implant Body Sens Networks, BSN 2017*. 2017;(May):197-201. doi:10.1109/BSN.2017.7936040

172. Davis I, Mulloy F. Midfoot Strikers Are Different from Forefoot Strikers, but Similar to Rearfoot Strikers. *Foot Ankle Orthop*. 2018;3(3):2473011418S0004. doi:10.1177/2473011418s00041

173. Zeng M, Nguyen LT, Yu B, et al. Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors. *Int Conf Mob Comput Appl Serv*. 2014. doi:10.4108/icst.mobicase.2014.257786

174. Zappi P, Lombriser C, Stiefmeier T, et al. Activity recognition from on-body sensors: Accuracy-power trade-off by dynamic sensor selection. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2008;4913 LNCS:17-33. doi:10.1007/978-3-540-77690-1_2

175. Roggen D, Calatroni A, Rossi Mi, et al. Collecting complex activity datasets in highly rich networked sensor environment.pdf. In: *IEEE Seventh International Conference on Networked Sensing Systems*. ; 2010.

176. Lockhart JW, Weiss GM, Xue JC, Gallagher ST, Grosner AB, Pulickal TT. Design considerations for the WISDM smart phone-based sensor mining architecture. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min*. 2011:25-33. doi:10.1145/2003653.2003656

177. Yoon H, Hwan S, Ho S, Choi J, Jin Y. Wakefulness evaluation during sleep for healthy subjects and OSA patients using a patch-type device. *Comput Methods Programs Biomed*. 2018;155:127-138.

178. Gordienko Y, Stirenko S, Kochura Y, Alienin O, Novotarskiy M, Gordienko N. Deep Learning for Fatigue Estimation on the Basis of Multimodal Human-Machine Interactions. *ArXiv*. 2017. http://arxiv.org/abs/1801.06048.

179. Caldwell JA, Caldwell JL, Brown DL, Smith JK. The Effects of 37 Hours of Continuous Wakefulness On the Physiological Arousal, Cognitive Performance, Self-Reported Mood, and Simulator Flight Performance of F-117A Pilots. *Mil Psychol*. 2004;16(3):163-181. doi:10.1207/s15327876mp1603_2

180. Chen J, Song X, Lin Z. Revealing the "invisible Gorilla" in construction: Estimating construction safety through mental workload assessment. *Autom Constr*. 2016;63:173-183. doi:10.1016/j.autcon.2015.12.018

181. Pickett J, Hofmans J. Stressors, Coping Mechanisms, and Uplifts of Commercial Fishing in Alaska: A Qualitative Approach to Factors Affecting Human Performance in Extreme Environments. *j Hum Perform Extrem Environements*. 2019;15(1).

182. Hart SG, Staveland LE. Development of NASA TLX (Task Load Index): Results of empirical and theoretical research. *Adv Psychol*. 1988;52(3437):139-183.

183. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: A New Instrument Psychiatric Practice and Research. *Psychiatry Res*. 1989;(28):193-213. doi:10.1016/0165-1781(89)90047-4

184. Oliveira PA, Araújo D, Abreu AM. Proneness for exercise, cognitive and psychophysiological consequences of action observation. *Psychol Sport Exerc*. 2014;15(1):39-47. doi:10.1016/j.psychsport.2013.09.008

185. Graybiel A, Wood CD, Miller II EF, et al. Diagnostic criteria for grading the severity of motion sickness. *Aerosp Med*. 1968;39:453-455.

186. Bajaj JS, Heuman DM, Sterling RK, et al. Validation of EncephalApp, Smartphone-Based Stroop Test, for the Diagnosis of Covert Hepatic Encephalopathy. *Clin Gastroenterol Hepatol*. 2015;13(10):1828-1835. doi:10.1016/j.cgh.2014.05.011

187. Johnstone JA, Ford PA, Hughes G, Watson T, Mitchell ACS, Garrett AT. Field based reliability and validity of the bioharness^{TM} multivariable monitoring device. *J Sports Sci Med*. 2012;11(4):643-652.

188. Wang J, Chen Y, Hao S, Peng X, Hu L. Deep learning for sensor-based activity recognition: A Survey. *Pattern Recognit Lett*. 2018;119:3-11. doi:10.1016/j.patrec.2018.02.010

189. Nweke HF, Teh YW, Mujtaba G, Al-garadi MA. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Inf Fusion*. 2019;46:147-170. doi:10.1016/j.inffus.2018.06.002

190. Dhiman C, Vishwakarma DK. A review of state-of-the-art techniques for abnormal human activity recognition. *Eng Appl Artif Intell*. 2019;77(August 2018):21-45. doi:10.1016/j.engappai.2018.08.014

191. Nweke HF, Teh YW, Al-garadi MA, Alo UR. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Syst Appl*. 2018;105:233-261.

doi:10.1016/j.eswa.2018.03.056

192. Chen Y, Xue Y. A deep learning approach to human activity recognition based on single accelerometer. *Proc - 2015 IEEE Int Conf Syst Man, Cybern SMC 2015*. 2016:1488-1492. doi:10.1109/SMC.2015.263

193. Williams D., Murray N., Douglas W. Athletes who train on unstable compared to stable surfaces exhibit unique postural contral strategies in response to balance pertubations. *J Sport Heal Sci*. 2016;5:70-76.

194. Wang G, Li Q, Wang L, Wang W, Wu M, Liu T. Impact of sliding window length in indoor human motion modes and pose pattern recognition based on smartphone sensors. *Sensors (Switzerland)*. 2018;18(6):1965. doi:10.3390/s18061965

195. Zheng X, Wang M, Ordieres-Meré J. Comparison of data preprocessing approaches for applying deep learning to human activity recognition in the context of industry 4.0. *Sensors (Switzerland)*. 2018;18(7):1-13. doi:10.3390/s18072146

196. Stein S, McKenna SJ. Recognising complex activities with histograms of relative tracklets. *Comput Vis Image Underst*. 2017;154:82-93. doi:10.1016/j.cviu.2016.08.012

197. Cust EE, Sweeting AJ, Ball K, Robertson S. Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. *J Sports Sci*. 2019;37(5):568-600. doi:10.1080/02640414.2018.1521769

198. Siirtola, Pekka; Laurinen, Perttu; Haapalainen, Eija; Roning J. Clustering-based activity classification with a wrist-worn accelerometer using basic features.pdf. *Comput Intell Data Mining, 2009 CIDM '09 IEEE Symp*. 2009:95-100. doi:10.1109/CIDM.2009.4938635

199. Johnson WR, Alderson JA, Lloyd DG, Mian A. Predicting athlete ground reaction forces and moments from spatio-temporal driven CNN models. *IEEE Trans Biomed Eng*. 2019;66(3):689-694.

200. Xu, Chai, He, Zhang, Duan. HAR A Deep Neural Network for Complex Human Activity Recognition. 2019.

201. Sarma N, Chakraborty S, Banerjee DS. Activity recognition through feature learning and annotations using LSTM. 2019;2061:444-447. doi:10.1109/comsnets.2019.8711147

202. Banos O, Galvez JM, Damas M, Pomares H, Rojas I. Window size impact in human activity recognition. *Sensors (Switzerland)*. 2014. doi:10.3390/s140406474

203. Kwapisz JR, Weiss GM, Moore SA. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explor Newsl*. 2011;12(2):74. doi:10.1145/1964897.1964918

204. Johnstone JA, Ford PA, Hughes G, Watson T, Garrett AT. Bioharness TM multivariable monitoring device. Part I: Validity. *J Sport Sci Med*. 2012;11(March):400-408. http://www.jssm.org.

205. Johnstone JA, Ford PA, Hughes G, Watson T, Garrett AT. Bioharness^{TM} multivariable monitoring device. Part II: Reliability. *J Sport Sci Med*. 2012;11(3):409-417.

206. Shigihara Y, Tanaka M, Ishii A, et al. Two different types of mental fatigue produce different styles of task performance. *Neurol Psychiatry Brain Res*. 2013;19(1):5-11. doi:10.1016/j.npbr.2012.07.002

207. Miyake S, Yamada S, Shoji T, Takae Y, Kuge N, Yamamura T. Physiological responses to workload change. A test/retest examination. *Appl Ergon*. 2009;40(6):987-996. doi:10.1016/j.apergo.2009.02.005

208. Kwapisz JR, Weiss GM, Moore SA. WISDM (Wireless Sensor Data Mining) Project. Fordham University, Department of Computer and Information Science, http://storm.cis.fordham.edu/~gweiss/ wisdm/. 2010. doi:10.1145/1964897.1964918

209. Phillips RO. A review of definitions of fatigue - And a step towards a whole definition. *Transp Res Part F Traffic Psychol Behav*. 2015;29:48-56. doi:10.1016/j.trf.2015.01.003

210. Vural E, Mujdat C AE. Machine Learning Systems for Detecting Driver Drowsiness. *Digit Signal Process in-Vehicle Mob Syst*. 2007;(June):16937–16953.

211. Desai A V., Haque MA. Vigilance monitoring for operator safety: A simulation study on highway driving. *J Safety Res*. 2006;37(2):139-147. doi:10.1016/j.jsr.2005.11.003

212. Garcés Correa A, Orosco L, Laciar E. Automatic detection of drowsiness in EEG records based on multimodal analysis. *Med Eng Phys*. 2014;36(2):244-249. doi:10.1016/j.medengphy.2013.07.011

213. Duncan MJ, Fowler N, George O, Joyce S, Hankey J. Mental Fatigue Negatively Influences Manual Dexterity and Anticipation Timing but not Repeated High-intensity Exercise Performance in Trained Adults. *Res Sport Med*. 2015;23(1):1-13. doi:10.1080/15438627.2014.975811

214. Park K, Rosengren KS, Horn GP, Smith DL, Hsiao-Wecksler ET. Assessing gait changes in firefighters due to fatigue and protective clothing. *Saf Sci*. 2011;49(5):719-726. doi:10.1016/j.ssci.2011.01.012

215. Smith BP, Browne M, Armstrong TA, Ferguson SA. The accuracy of subjective measures for assessing fatigue related decrements in multi-stressor environments. *Saf Sci*. 2016;86:238-244. doi:10.1016/j.ssci.2016.03.006

216. Drew Dawson, Kathryn Reid. Fatigue alcohol and performance impairment. *Nature*. 1997;388(1):235-237.

217. Davis MP, Walsh D. Mechanisms of Fatigue. *J Support Oncol J Support Oncol*. 2010;8(November):164-174. doi:10.1016/B978-0-443-07427-1.50014-5

218. Hampson DB, Gibson AS, Lambert MI, Noakes TD. The Influence of Sensory Cues on the Perception of Exertion During Exercise and Central Regulation of Exercise Performance. *Sport Med*. 2001;31(13):935-952. doi:10.2165/00007256-200131130-00004

219. Noakes TD. Physiological models to understand exercise fatigue and the adaptations that predict or enhance athletic performance. *Med Sci Sport*. 2000:123-145.

220. Van Cutsem J, Marcora S, De Pauw K, Bailey S, Meeusen R, Roelands B. The Effects of Mental Fatigue on Physical Performance: A Systematic Review. *Sport Med*. 2017;47(8):1569-1588. doi:10.1007/s40279-016-0672-0

221. Borg G, Borg E. To determine the magnitude of pain with Borg. *Fechner Day 2014 – Proc 30th Annu Meet Int Soc Psychophys*. 2014;45:16.

222. Millet GY, Tomazin K, Verges S, et al. Neuromuscular consequences of an extreme mountain ultra-marathon. *PLoS One*. 2011;6(2):e17059. doi:10.1371/journal.pone.0017059

223. Portenga ST, Aoyagi MW, Cohen AB. Helping to build a profession: A working definition of sport and performance psychology. *J Sport Psychol Action*. 2017;8(1):47-59. doi:10.1080/21520704.2016.1227413

224. Lambert E V, St Clair Gibson A, Noakes TD. Complex systems model of fatigue: integrative homoeostatic control of peripheral physiological systems during

exercise in humans. *Br J Sports Med*. 2004;39(1):52-62. doi:10.1136/bjsm.2003.011247

225. Jo J, Lee SJ, Park KR, Kim IJ, Kim J. Detecting driver drowsiness using feature-level fusion and user-specific classification. *Expert Syst Appl*. 2014;41(4 PART 1):1139-1152. doi:10.1016/j.eswa.2013.07.108
226. Zhu Y, Jankay RR, Pieratt LC, Mehta RK. Wearable sensors and their metrics for measuring comprehensive occupational fatigue: A scoping review. *Proc Hum Factors Ergon Soc*. 2017;2017-Octob:1041-1045. doi:10.1177/1541931213601744
227. St Clair Gibson A, Goedecke JH, Harley YX, et al. Metabolic setpoint control mechanisms in different physiological systems at rest and during exercise. *J Theor Biol*. 2005;236(1):60-72. doi:10.1016/j.jtbi.2005.02.016
228. Wickens CD, Keller JW, Shaw C. Human Factors in High-Altitude Mountaineering. *J Hum Perform Extrem Environ*. 2015;12(1):5-8. doi:10.7771/2327-2937.1065
229. Apple. Research kit. https://developer.apple.com/researchkit/.
230. Kovářová L, Pánek D, Kovář K, Hlinčík Z. Relationship between subjectively perceived exertion and objective loading in trained athletes and non-athletes. *J Phys Educ Sport*. 2015;15(2):186-193. doi:10.7752/jpes.2015.02029
231. Winter DA. Kinematic and kinetic patterns in human gait: variability and compensating effects. *Hum Mov Sci 3, 51-76.ment Sci 3, 51-76*. 1984;3:51-76.
232. St Clair Gibson A, Lambert E V, Rauch LHG, et al. The role of information processing between the brain and peripheral physiological systems in pacing and perception of effort. *Sport Med*. 2006;36(8):705-722. doi:10.2165/00007256-200636080-00006
233. Noakes TD. Time to move beyond a brainless exercise physiology: the evidence for complex regulation of human exercise performance. *Appl Physiol Nutr Metab*. 2011;36(1):23-35. doi:10.1139/H10-082
234. Enoka RM, Duchateau J. *Translating Fatigue to Human Performance*. Vol 48.; 2016. doi:10.1249/MSS.0000000000000929
235. Enoka RM. Mechanisms of muscle fatigue: Central factors and task dependency. *J Electromyogr Kinesiol*. 1995;5(3):141-149. doi:10.1016/1050-6411(95)00010-W
236. Wu Y, Chen P, Luo X, et al. Measuring signal fluctuations in gait rhythm time

series of patients with Parkinson's disease using entropy parameters. *Biomed Signal Process Control*. 2017;31:265-271. doi:10.1016/j.bspc.2016.08.022

237. Zwarts MJ, Bleijenberg G, van Engelen BGM. Clinical neurophysiology of fatigue. *Clin Neurophysiol*. 2008;119(1):2-10. doi:10.1016/j.clinph.2007.09.126
238. Pereira HM, Keller ML. Understanding the mechanisms of neuromuscular fatigue with paired-pulse stimulation. *J Physiol*. 2012;590(1):5-6. doi:10.1113/jphysiol.2011.219204
239. Clermont C, Pohl A, Ferber R. Clermont 2019 Fatigue-Related Changes in Running Gait Patterns Persist in the Days Following a Marathon Race.pdf. *J od Sport Rehabil*. 2019:1-8.
240. Straus E, Sherman E, Spreen O. *A Compendium of Neuropsychological Tests*. Third. (Oxford University Press, ed.).; 2006.
241. Liu J, Zhang C, Zheng C. EEG-based estimation of mental fatigue by using KPCA-HMM and complexity parameters. *Biomed Signal Process Control*. 2010;5(2):124-130. doi:10.1016/j.bspc.2010.01.001
242. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986;323(6088):533-536. doi:10.1038/323533a0
243. Tobergte DR, Curtis S. Improving Neural Networks with Dropout. *J Chem Inf Model*. 2013;53(9):1689-1699. doi:10.1017/CBO9781107415324.004
244. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J Mach Learn Res*. 2014;15:1929-1958. doi:10.1214/12-AOS1000
245. Noakes TD, Peltonen JE, Rusko HK. Evidence that a central governor regulates exercise performance during acute hypoxia and hyperoxia. *J Exp Biol*. 2001;204(18):3225-3234.
246. Peters A, McEwen BS, Friston K. Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. *Prog Neurobiol*. 2017;156:164-188. doi:10.1016/j.pneurobio.2017.05.004
247. Marcora SM, Staiano W, Manning V. Mental fatigue impairs physical performance in humans. *J Appl Physiol*. 2009;106(3):857-864. doi:10.1152/japplphysiol.91324.2008
248. Nindl BC. State of the Science of Military Human Performance Optimization. *State Sci Symp Ser Fit Heal Outcomes Exerc Heal Nutr Wounded, Inj Ill Veterans*. 2016;(March).

249. Cairns SP. The central governor of exercise performance: fact or fiction. *New Zeal J Sport Med*. 2011;38(2):47-50. doi:10.1093/carcin/bgr169

APPENDIX A. ETHICS FORMS FOR EXPERIMENT 1 - OFFSHORE SAILING

AUTEC Secretariat

Auckland University of Technology
D-88, WU406 Level 4 WU Building City Campus
T: +64 9 921 9999 ext. 8316
E: ethics@aut.ac.nz
www.aut.ac.nz/researchethics

1 December 2017

Patria Hume
Faculty of Health and Environmental Sciences

Dear Patria

Re Ethics Application: **17/353 Measurement of cognitive and physical performance during off shore sailing**

Thank you for providing evidence as requested.

Your ethics application has been approved for three years until 1 December 2020.

It is noted that data was collected and consented for prior to final ethics approval.

Standard Conditions of Approval

1. A progress report is due annually on the anniversary of the approval date, using form EA2, which is available online through http://www.aut.ac.nz/researchethics.
2. A final report is due at the expiration of the approval period, or, upon completion of project, using form EA3, which is available online through http://www.aut.ac.nz/researchethics.
3. Any amendments to the project must be approved by AUTEC prior to being implemented. Amendments can be requested using the EA2 form: http://www.aut.ac.nz/researchethics.
4. Any serious or unexpected adverse events must be reported to AUTEC Secretariat as a matter of priority.
5. Any unforeseen events that might affect continued ethical acceptability of the project should also be reported to the AUTEC Secretariat as a matter of priority.

Please quote the application number and title on all future correspondence related to this project.

AUTEC grants ethical approval only. If you require management approval for access for your research from another institution or organisation then you are responsible for obtaining it. If the research is undertaken outside New Zealand, you need to meet all locality legal and ethical obligations and requirements. You are reminded that it is your responsibility to ensure that the spelling and grammar of documents being provided to participants or external organisations is of a high standard.

For any enquiries, please contact ethics@aut.ac.nz

Yours sincerely,

Kate O'Connor
Executive Manager
Auckland University of Technology Ethics Committee

Cc: bkrussell10@gmail.com

Participant Information Sheet

Date Information Sheet Produced: 26 Sept 2017

PROJECT TITLE

Fatigue assessment during offshore sailing event

During Event

- Sensor - Participants will wear a BioHarness. Devices will be charged for 1 hour – not worn – during lunch 12:00 local time. (Zephyr/Medtronic - https://www.zephyranywhere.com/users/academic-researchers). Heart Rate, Respiration, HRV and 3 axis accelerometer data will be collected.
- Diary - 4 times a day; waking, just prior to sleep, 10am, 4pm a test battery will include. See Attached Diary Spreadsheet (4 Diary Fatigue Hume …xls).
    1. Finger Tap Test
    2. Stroop Test
    3. Weather, sea state (height and direction of waves) and yacht speed will also be recorded in the ships log and a copy provided at the end of the trip. These numbers directly effect the environment the crew is exposed to; anxiety, comfort, g forces, sleep disturbances.

AN INVITATION

You are invited to take part in the above mentioned research project. Your participation in this research is voluntary. You are free to withdraw consent and discontinue participation from the study at any without influencing any present and/or future involvement with the Auckland University of Technology.

Your consent to participate in this research will be indicated by your signing and dating the consent form. Signing the consent form indicates that you have freely given your consent to participate, and that there has been no coercion or inducement to participate by the researchers from AUT.

WHAT IS THE PURPOSE OF THIS RESEARCH?

The purpose of the study is to determine vital signs and movement to identify levels of cognitive and physical fatigue. Previously validated assessment tests on an iPhone will be used as the control. Both sets of data will be used to train a neural network to assess fatigue in the field without the need to answer questions.

HOW WAS I CHOSEN TO BE ASKED TO PARTICIPATE IN THE RESEARCH?

You were chosen to participate in the study as you're are taking part in the event.

HOW DO I AGREE TO PARTICIPATE IN THIS RESEARCH?

Your participation in this research is voluntary (i.e. it is your choice) and whether or not you choose to participate will neither advantage nor disadvantage you. You are able to withdraw from the study at any time. If you choose to withdraw from the study, then you will be offered the choice between having any data that is identifiable as belonging to you removed or allowing it to continue to be used. However, once the findings have been produced, removal of your data may not be possible.

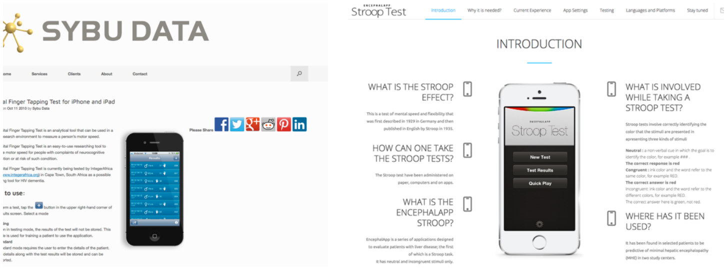
WHAT HAPPENS IN THIS RESEARCH?

You will be asked to complete a medical questionnaire that will be recorded and used as part of the analysis of the data obtained. This history will be recorded in a database only accessable to the investigators directly involved in the study.

During the event you will wear the BioHarness and perform a few tests on an iPhone each day. The BioHarness measures vital signs and the tests measure cognitive fatigue.The first test is the Finger Tap Test, FTT, where you will be guided to tap the screen as fast as possible for a minute. Another test is called the Stroop Test, where there will be words green, blue red which may be in opposing colours e.g. the word “blue” in red font. You will be asked to tap a button for each colour based on the font color. Studies have shown both these tests changed with sleep and other fatigueing situations. On a daily dairy you will be asked to answer some questions on hwo you are feeling.

The BioHarness is a fabric chest strap that can measure your heart rate, breathing rate and movement. It will need to be charged for an hour a day, typically over lunch time.

Finger Tap Test and Stroop Test



al Finger Tapping Test for iPhone and iPad
n Oct 11 2010 by Sybu Data

ital Finger Tapping Test is an analytical tool that can be used in a esearch environment to measure a person's motor speed.

ital Finger Tapping Test is an easy-to-use researching tool to e motor speed for people with complaints of neurocognitive tion or at risk of such condition.

ital Finger Tapping Test is currently being tested by IntegerAfrica ww.integerafrica.org) in Cape Town, South Africa as a possible g tool for HIV dementia.

to use:

rm a test, tap the + button in the upper right-hand corner of ults screen. Select a mode

ing
n in testing mode, the results of the test will not be stored. This e is used for training a patient to use the application.
dard
dard mode requires the user to enter the details of the patient. details along with the test results will be stored and can be rted.

INTRODUCTION

WHAT IS THE STROOP EFFECT?

This is a test of mental speed and flexibility that was first described in 1929 in Germany and then published in English by Stroop in 1935.

HOW CAN ONE TAKE THE STROOP TESTS?

The Stroop test have been administered on paper, computers and on apps.

WHAT IS THE ENCEPHALAPP STROOP?

EncephalApp is a series of applications designed to evaluate patients with liver disease; the first of which is a Stroop task.
It has neutral and incongruent stimuli only.

WHAT IS INVOLVED WHILE TAKING A STROOP TEST?

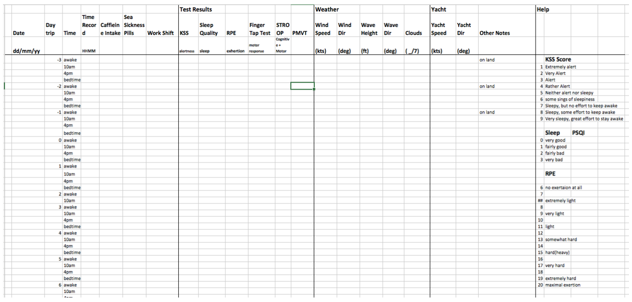
Stroop tests involve correctly identifying the color that the stimuli are presented in apresenting three kinds of stimuli

Neutral : a non-verbal cue in which the goal is to identify the color, for example ### .
The correct response is red
Congruent : ink color and the word refer to the same color, for example RED.
The correct answer is red
Incongruent: ink color and the word refer to the different colors, for example RED.
The correct answer here is green, not red.

WHERE HAS IT BEEN USED?

It has been found in selected patients to be predictive of minimal hepatic encephalopathy (MHE) in two study centers.

Results will be recorded in the app by allocating each person an ID and also manually recording the results on the daily diary.



							Test Results						Weather					Yacht			Help		
Date	Day trip	Time	Time Record	Caffieine Intake	Sea Sickness Pills	Work Shift	KSS	Sleep Quality	RPE	Finger Tap Test	STROOP	PMVT	Wind Speed	Wind Dir	Wave Height	Wave Dir	Clouds	Yacht Speed	Yacht Dir	Other Notes			
dd/mm/yy			HHMM				alertness	sleep	exhertion	motor response	Cognitive + Motor		(kts)	(deg)	(ft)	(deg)	(_/7)	(kts)	(deg)				
	-3	awake																	on land		KSS Score	
		10am																		1	Extremely alert	
		4pm																		2	Very Alert	
		bedtime																		3	Alert	
	-2	awake																	on land	4	Rather Alert	
		10am																		5	Neither alert nor sleepy	
		4pm																		6	some sings of sleepiness	
		bedtime																		7	Sleepy, but no effort to keep awake	
	-1	awake																	on land	8	Sleepy, some effort to keep awake	
		10am																		9	Very sleepy, great effort to stay awake	
		4pm																				
		bedtime																			Sleep	PSQI
	0	awake																		0	very good	
		10am																		1	fairly good	
		4pm																		2	fairly bad	
		bedtime																		3	very bad	
	1	awake																				
		10am																			RPE	
		4pm																				
		bedtime																		6	no exertaion at all	
	2	awake																		7		
		10am																		##	extremely light	
	3	awake																		8		
		10am																		9	very light	
		4pm																		10		
		bedtime																		11	light	
	4	awake																		12		
		10am																		13	somewhat hard	
		4pm																		14		
		bedtime																		15	hard(heavy)	
	5	awake																		16		
		10am																		17	very hard	
		4pm																		18		
		bedtime																		19	extremely hard	
	6	awake																		20	maximal exertion	
		10am																				

The study duration will be the sailiing trip from the Cheasapeake to the Caribean. Only fatigue will be studied over the trip.

WHAT ARE THE DISCOMFORTS AND RISKS?

Risks are only those discomforts and risks that normally occur from participating. The BioHarness is worn around the chest. Assistance and training will be offered to maximise fit and comport. It will not constrict breathing and should not be noticeable after a few minutes of wearing it.

WHAT ARE THE BENEFITS?

Information gained from this research has the potential to assist operators and sport people in the field to assess fatigue real time and plan appropriate strategies for success and minimise risk.

WHAT COMPENSATION IS AVAILABLE FOR INJURY OR NEGLIGENCE?

There is no compensation for this research and you are undertaking the activity voluntarily. It is your responsibility to always operate safely and not allow any unforeseen consequences of the data collection to influence the safe working practices.

HOW IS MY PRIVACY PROTECTED?

The data from the project will be coded and held anonymously in secure storage under the responsibility of the principal investigator of the study in accordance with the requirements of the New Zealand Privacy Act (1993).

All reference to participants will be by code number only in terms of the research project and publications. Identification information will be stored on a separate file and computer from that containing the actual data.

Only the investigators will have access to computerised data obtained from the participants.

In the case of you recording a potential medical issue it will be discussed with the appropriate person or guardian.

WHAT ARE THE COSTS OF PARTICIPATING?

There is no additional costs involved with this research, all equipment is supplied.

OPPORTUNITY TO CONSIDER INVITATION

Please take the necessary time you need to consider the invitation to participate in this research. It is reiterated that your participation in this research is completely voluntary.

If you require further information about the research topic please feel free to contact Brian Russell (details are at the bottom of this information sheet).

You may withdraw from the study at any time without there being any adverse consequences of any kind.

You may ask for a copy of your results at any time and you have the option of requesting a report of the research outcomes at the completion of the study.

HOW DO I JOIN THE STUDY?

If you are interested in participating in this research please feel free to contact Brian Russell (details are at the bottom of this information sheet).

PARTICIPANT CONCERNS

Any concerns regarding the nature of this project should be notified in the first instance to the Project Supervisor, Professor Patria Hume. Email: phume@aut.ac.nz or phone +64 9 921 9999 ext. 7306.

Concerns regarding the conduct of the research should be notified to the Executive Secretary of AUTEC, Kate O'Connor, ethics@aut.ac.nz, (09) 921 9999 ext. 6038.

Whom do I contact for further information about this research?

Please keep this Information Sheet and a copy of the Consent Form for your future reference. You are also able to contact the research team as follows:

Researchers Contact Details:

Brian Russell, email: bkrussell10@gmail.com, phone +64 22 378 6093

Project Supervisor Contact Details

Professor Patria Hume, Sports Performance Research Institute New Zealand, School of Sport and Recreation, Auckland University of Technology. Email: phume@aut.ac.nz or phone +64 9 921 9999 ext. 7306.

Approved by the Auckland University of Technology Ethics Committee on *type the date final ethics approval was granted*, AUTEC Reference number *type the reference number*

AUT SPORTS PERFORMANCE
RESEARCH INSTITUTE NEW ZEALAND

PARTICIPANT MEDICAL HISTORY

This medical form is to identify any illnesses or medical history that may affect the protocols or outcomes for the research. It will not be communicated to any party other than the research team and will be stored under AUTEL guidelines for patient data. Data will be anonymized.

Team:	Role(if applicable):	Age: in yrs
Name:	Height: cms Weight: kgs	BMI =

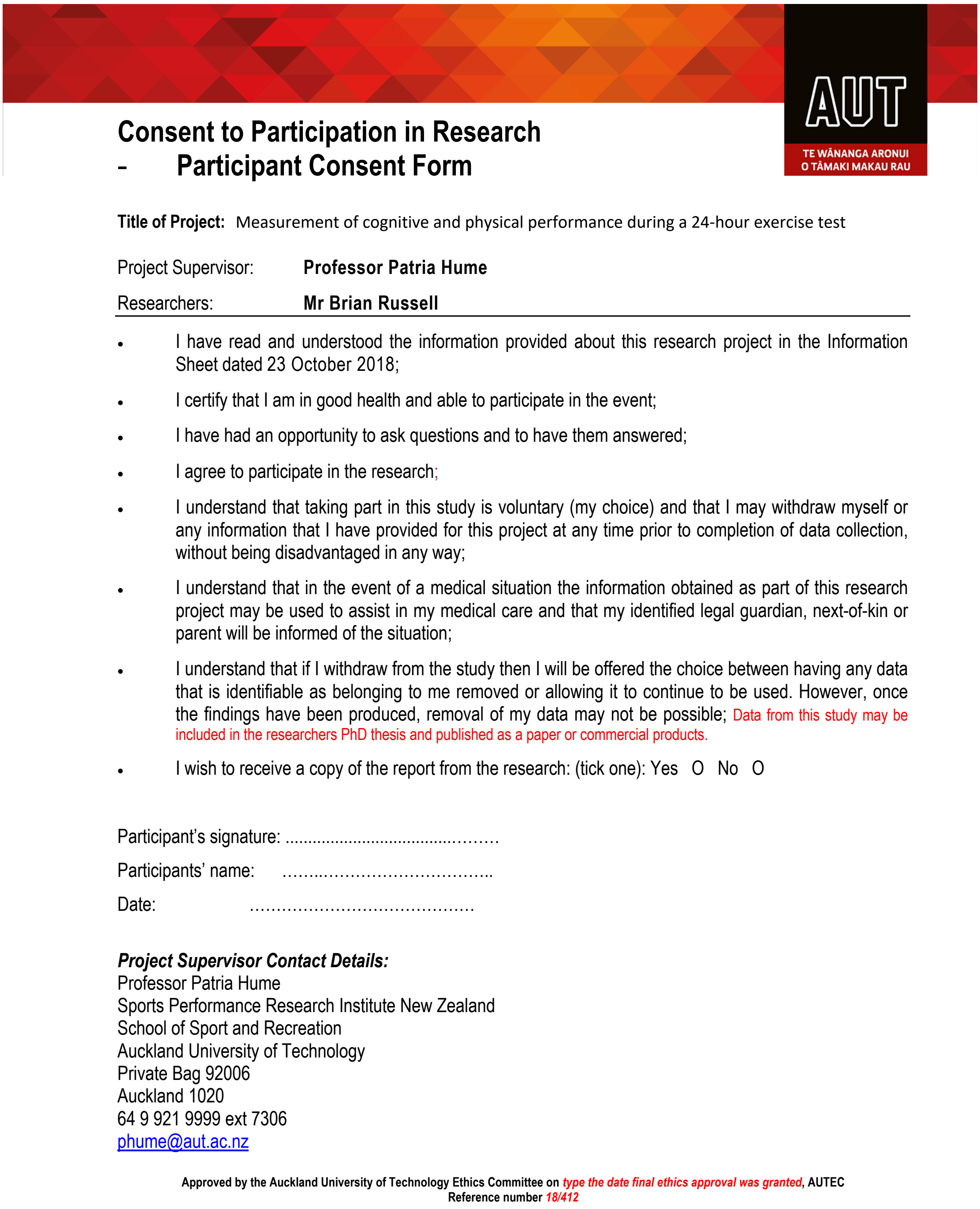
Previous Medical History or Illness

Have you ever had or do you now have:					
	Yes	No		Yes	No
Arrhythmia	☐	☐	Heart Problems	☐	☐
Balance issues or Vertigo	☐	☐	Breathing Problems including Asthma/Bronchitis/Emphysema		
Brain Injuries (not concussion)	☐	☐	High/Low Blood Pressure	☐	☐
Brain Surgery			Ulcers		
Chest Pains	☐	☐	Bowel Problems	☐	☐
Clinical photosensitivity	☐	☐	Urinary Infections	☐	☐
Concussion	☐	☐	Kidney Problems	☐	☐
Deafness/Ear Problems	☐	☐	Diabetes	☐	☐
Eye Problems	☐	☐	Blood transfusions	☐	☐
Faints, Fits or 'Funny turns'	☐	☐	Hepatitis	☐	☐
Glasses/Contacts	☐	☐	Thyroid disorder	☐	☐
			Allergies	☐	☐
High levels of stress	☐	☐	(Specify):		
Typical coffee per day – week day		Cups	Fractures:	☐	☐
Typical coffee per day - weekend		Cups	Specify		
Typical hours of sleep per day		Hours	Sprains	☐	☐
Typical hours per week of moderate exercise (e.g. walking)		hours	Specify:		
Typical hours per week of high intensity exercise (e.g. running, cardio class etc)		Hours			
Do typically you eat breakfast	☐	☐	Operations:	☐	☐
Do typically you eat lunch	☐	☐	Specify:		
Do typically you eat dinner	☐	☐			
			Infectious diseases	☐	☐

LIST ANY OTHER HEALTH PROBLEMS NOT IDENTIFIED ABOVE/RELEVANT INFORMATION FOR THOSE TICKED YES

PLEASE LIST ALL MEDICATIONS YOU ARE CURRENTLY USING
Prescribed:
Non prescribed / Over the counter / Complimentary:

APPENDIX B. ETHICS FORMS FOR EXPERIMENT 2 – TRAIL RUN IN THE MOUNTAINS



Consent to Participation in Research

- Participant Consent Form

Title of Project: Measurement of cognitive and physical performance during a 24-hour exercise test

Project Supervisor: **Professor Patria Hume**

Researchers: **Mr Brian Russell**

- I have read and understood the information provided about this research project in the Information Sheet dated 23 October 2018;
- I certify that I am in good health and able to participate in the event;
- I have had an opportunity to ask questions and to have them answered;
- I agree to participate in the research;
- I understand that taking part in this study is voluntary (my choice) and that I may withdraw myself or any information that I have provided for this project at any time prior to completion of data collection, without being disadvantaged in any way;
- I understand that in the event of a medical situation the information obtained as part of this research project may be used to assist in my medical care and that my identified legal guardian, next-of-kin or parent will be informed of the situation;
- I understand that if I withdraw from the study then I will be offered the choice between having any data that is identifiable as belonging to me removed or allowing it to continue to be used. However, once the findings have been produced, removal of my data may not be possible; Data from this study may be included in the researchers PhD thesis and published as a paper or commercial products.
- I wish to receive a copy of the report from the research: (tick one): Yes O No O

Participant's signature: ....……

Participants' name: ……..………………………..

Date: ……………………………………

Project Supervisor Contact Details:

Professor Patria Hume
Sports Performance Research Institute New Zealand
School of Sport and Recreation
Auckland University of Technology
Private Bag 92006
Auckland 1020
64 9 921 9999 ext 7306
phume@aut.ac.nz

Approved by the Auckland University of Technology Ethics Committee on *type the date final ethics approval was granted*, AUTEC Reference number *18/412*

Participant Information Sheet

Date Information Sheet Produced: 23 October 2018

PROJECT TITLE

Measurement of cognitive and physical performance during a 24-hour exercise test

AN INVITATION

You are invited to take part in the above mentioned research project. Your participation in this research is voluntary. You are free to withdraw consent and discontinue participation from the study at any without influencing any present and/or future involvement with the Auckland University of Technology.

Your consent to participate in this research will be indicated by your signing and dating the consent form. Signing the consent form indicates that you have freely given your consent to participate, and that there has been no coercion or inducement to participate by the researchers from AUT.

WHAT IS THE PURPOSE OF THIS RESEARCH?

The purpose of the study is to determine vital signs and movement to identify levels of cognitive and physical fatigue in order to allow an AI model to be built allow a sensor to replace questionnaires and other testing in the field. An iPhone will be used as the control to run an app to test cognitive and physical fatigue. Both sets of data will be used to train a neural network to assess fatigue in the field without the need to answer questions.

HOW WAS I CHOSEN TO BE ASKED TO PARTICIPATE IN THE RESEARCH?

You were chosen to participate in the study as you're have an average fitness and an interest in fitness and cognitive performance.

HOW DO I AGREE TO PARTICIPATE IN THIS RESEARCH?

Your participation in this research is voluntary (i.e. it is your choice) and whether or not you choose to participate will neither advantage nor disadvantage you. You are able to withdraw from the study at any time. If you choose to withdraw from the study, then you will be offered the choice between having any data that is identifiable as belonging to you removed or allowing it to continue to be used. However, once the findings have been produced, removal of your data may not be possible.

WHAT HAPPENS IN THIS RESEARCH?

You will be asked to complete a medical questionnaire that will be recorded and used as part of the analysis of the data obtained. This history will be recorded in a database only accessable to the investigators directly involved in the study.

During the activity you will wear the BioHarness and perform a few activities every hour for a 24 hour period. Activities inlcude using an an iPhone app, running 2km and performing a computer program called a "Multi Attribute Test Battery" which which will ask you to carry out tasks in paralle. Once per day for 3 days prior and 3 days after the event you will perform the tests on the iPhone app. The BioHarness measures vital signs and the tests measure cognitive fatigue.The iPhone app will ask you some questions about sleepiness and motivation and then ask you to carry out some tasks such as tapping the screen as many times in 10 seconds, choosing colours, joining sequences of numbers and letters and doing mental addition.

The BioHarness is a fabric chest strap that can measure your heart rate, breathing rate and movement. It will need to be charged for an hour a day, typically over lunch time.

Finger Tap Test and Stroop Test

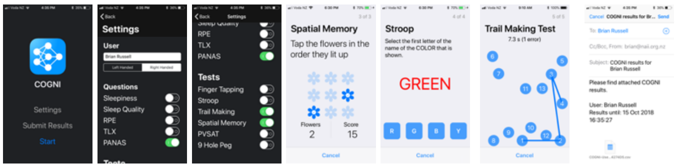


Figure 1. COGNI assessment app example screen shots

Multi Attribute Test Battery (MATB) for 15 minutes

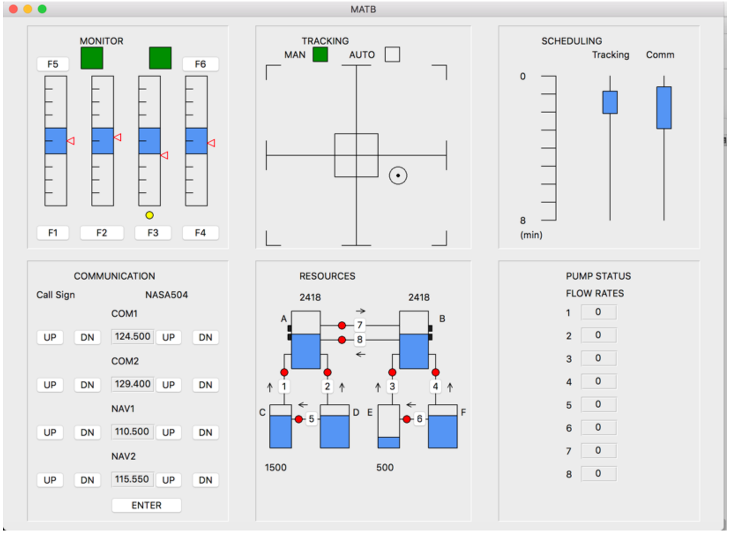
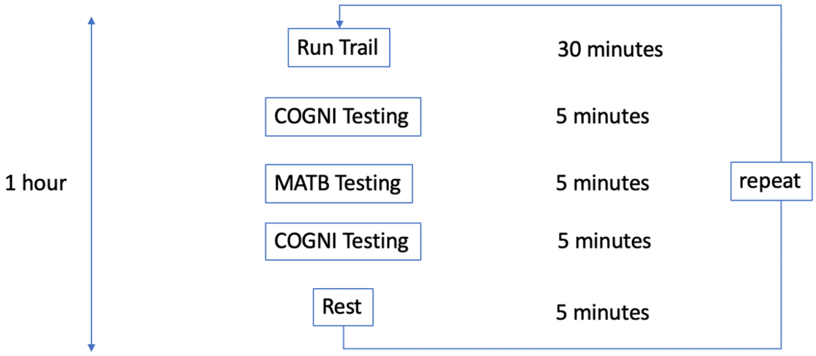


Figure 2. Multi Attribute Test Battery, coded in Python by the researcher. Based on Comstock 1992. The multi-Attribute Task Battery for Human Operator Workload and Strategic Behaviour Research.

Running will be a combination of suburban and trail running up a 150m high hill.

There is, preferably, no sleep during the 24 hour test period with no caffiene, nirmal hydration and nutrition. For 3 days either side of the activity day you are asked to have normal sleep, exercise and nutrition with no caffiene or alcohol.



WHAT LEVEL OF FITNESS IS REQUIRED?

It is anticipated you are of average fitness and run regularly.

WHAT ARE THE DISCOMFORTS AND RISKS?

The BioHarness is worn around the chest. Assistance and training will be offered to maximise fit and comport. It will not constrict breathing and should not be noticeable after a few minutes of wearing it.

WHAT ARE THE BENEFITS?

Information gained from this research has the potential to assist sports people and operators to assess fatigue real time and plan appropriate strategies for success and minimise risk in dangerous environments. Data from this study may be included in the researchers PhD thesis and published as a paper or commercial products.

WHAT COMPENSATION IS AVAILABLE FOR INJURY OR NEGLIGENCE?

There is no compensation for this research, and you are undertaking the activity voluntarily. It is your responsibility to always operate safely and not allow any unforeseen consequences of the data collection to influence the safe working practices. Travel costs may be covered.

HOW IS MY PRIVACY PROTECTED?

The data from the project will be coded and held anonymously in secure storage under the responsibility of the principal investigator of the study in accordance with the requirements of the New Zealand Privacy Act (1993).

All reference to participants will be by code number only in terms of the research project and publications. Identification information will be stored on a separate file and computer from that containing the actual data.

Only the investigators will have access to computerised data obtained from the participants.

In the case of you recording a potential medical issue it will be discussed with the appropriate person or guardian.

WHAT ARE THE COSTS OF PARTICIPATING?

There is no additional costs involved with this research, all equipment is supplied.

OPPORTUNITY TO CONSIDER INVITATION

Please take the necessary time you need to consider the invitation to participate in this research. It is reiterated that your participation in this research is completely voluntary.

If you require further information about the research topic please feel free to contact Brian Russell (details are at the bottom of this information sheet).

You may withdraw from the study at any time without there being any adverse consequences of any kind.

You may ask for a copy of your results at any time and you have the option of requesting a report of the research outcomes at the completion of the study.

HOW DO I JOIN THE STUDY?

If you are interested in participating in this research please feel free to contact Brian Russell (details are at the bottom of this information sheet).

PARTICIPANT CONCERNS

Any concerns regarding the nature of this project should be notified in the first instance to the Project Supervisor, Professor Patria Hume. Email: phume@aut.ac.nz or phone +64 9 921 9999 ext. 7306.

Concerns regarding the conduct of the research should be notified to the Executive Secretary of AUTEC, Kate O'Connor, ethics@aut.ac.nz, (09) 921 9999 ext. 6038.

Whom do I contact for further information about this research?

Please keep this Information Sheet and a copy of the Consent Form for your future reference. You are also able to contact the research team as follows:

Researchers Contact Details:

Brian Russell, email: bkrussell10@gmail.com, phone +64 22 378 6093

Project Supervisor Contact Details

Professor Patria Hume, Sports Performance Research Institute New Zealand, School of Sport and Recreation, Auckland University of Technology. Email: phume@aut.ac.nz or phone +64 9 921 9999 ext. 7306.

Approved by the Auckland University of Technology Ethics Committee on *type the date final ethics approval was granted*, AUTEC Reference number 18/412

APPENDIX C. FATIGUE AND VITAL SIGN MONITORING FOR OFFSHORE SAILING CREWS

This conference paper was from the first experiment in this PhD and lead to significant understanding of the challenges in collecting data from individuals in the field under fatiguing conditions and the challenge of validation.

Russell B.K.[1], Hume P.A.[1,2], McDaid A.[3], Simms, S.[4], ***Fatigue and Vital Sign Monitoring for Offshore Sailing Crews,*** NZ Journal of Sports Medicine, Nelson, NZ, October 2018 (abstract)

[1]Sports Performance Research Institute New Zealand (SPRINZ), Auckland University of Technology (AUT), Auckland;
[2]National Institute of Stoke and Applied Neurosciences (NISAN), AUT, Auckland;
[3]School of Engineering, University of Auckland;
[4]The University of Waikato, Hamilton.

Introduction: Cognitive and physical fatigue in multi-day adventure sport is key to performance and safety. Real world testing in the field can incorporate responses to risk and other factors that cannot be simulated in the lab. However field based research brings challenges in data collection, validation and practical considerations on support and compliance to the protocol in the presence of work load, danger and fatigue. Monitoring crews while off shore sailing with sensors is challenging as the crew is under stress and confined. Jetlag for international crew and sea sickness adds to the challenges. Multi day cognitive and physical assessment requires a protocol that can be adjusted for work shifts and circadian rhythms. *Aim:* To determine challenges of multi-day in-field research using tablet based assessments, questionnaires and sensor data for a four-person crew over a 12-day offshore sailing passage. *Methods:* The Stroop test (Egner & Hirsch, 2005) was used as it has been shown, using functional magnetic resonance imaging (fMRI) testing, to "stimulate left middle front gyrus (GFm) and superior frontal gyrus (GFs), and decreased activity in the bilateral prefrontal and partial cortices" which are indicated in high level perception and motor processes. The Finger Tap Test (FTT) was used for neurological (Amer et al., 2012), motor skills and neuromuscular fatigue measurement (Leyla & Kiziltan, 2016). The Borg Scale for Rating of Perceived Exertion (Borg, 1982) was used to indicate influence of central factors such as heart rate and peripheral factors such as blood lactate. The Borg scale

of 16 approximates a heart rate of 160 bpm.). A daily test battery of questions and tests was performed along with prolonged daily donning of a BioHarness (Medtronic MDT, formerly Zephyr) sensor with periods of non-use for charging and data down load. The BioHarness sensor was selected as it was comfortable for long term use by the crew and accurately measures vital signs and has an Inertial Measurement Unit (IMU). The test battery included questions asked and recorded to Excel (Microsoft Corp MSFT) and an iPhone (Apple Inc AAPL) with a different application for each test. Post analysis consisted of concatenating daily files, aligning date and time between sensor and questionnaires, and analysing vital sign data for features such as variability and entropy. A time based windowing analysis was performed to determine the best length for comparison to questionnaire data points. Sensor features were time aligned to questionnaire data. Correlation analysis between sensor features and questionnaires was performed.

Results: Compliance to the test and sensor protocol was not sufficient for a generalisable model between sensor and questionnaires. The questionnaires were performed by different applications which lead to undue effort for participants and researchers. Some data were lost during the upload process. The study showed a sensor should have sufficient data storage and battery life time for a multiple day event without depending on multi-step uploading and erasing. Vital sign variation and features did not correlate with previously validated cognitive assessments. Each individual adapted to the environment differently, negating any generalisations across the test population. Multi crew events in the field cannot assume subjects adapt in a similar manner and carry out tasks at the same time of day (e.g. sleep cycles , work shifts and circadian rhythms may be misaligned). Assessments need to be regular to capture time sensitivity and variation between subjects.

Conclusions: In multiday field events, research devices and protocols require a simplicity and reliability above that in the lab due to two main factors; there is no support in the field and the subjects and researcher are cognitively fatigued as part of the mission. Cognitive tests and questionnaires need to be hourly to provide sufficient information for correlation to sensor data and capture inter subject differences. Baselining before and after the event is required to separate fatigue effects from training and recovery effects. Cognitive tests should be on a single device with single application, with simple user input and minimisation of redundant steps between sessions, such as single login and only entering demographic data once for the entire research period. Inter-subject testing should minimise the likelihood of cross subject test data contamination. Sensors should be wearable for the entire event with no recharging, donning/doffing or downloading/erasing required. Battery management should ideally not require recharging of tablets and sensors for the entire multiday event. A remote display should be accessible to confirm vital signs are being correctly monitored.

APPENDIX D. A GENERATIVE MACHINE LEARNING FRAMEWORK FOR SYNTHESIZING SYMPTOMATIC ECG ASTRONAUT HEALTH DATA

This was part of the SETI NASA Foundation Development Lab summer program. I mentored the four researchers and contributed to the data collection, modelling approach and analysis. The first four authors were the group of researchers carrying out the main programming and paper writing.

Paper Accepted by NASA IWS Conference

Eleni Antoniadou*[1], David Belo*[1], Krittika D'Silva*[1], Brian Wang*[1], Brian Russell[1], Frank Soboczenski[1], Annie Martin[2], Graham Mackintosh[1,3], and Tianna Shaw[3]
[1] NASA Frontier Development Lab,
[2] Canadian Space Agency,
[3] NASA Ames Research Center

INTRODUCTION: Future interplanetary manned missions require safe and reliable medical monitoring and diagnostics due to the pathologies caused by radiation and micro-gravity [1]. Different medical modalities, such as wearable devices or point-of-care diagnostics, exist in spacecrafts to facilitate health monitoring, but early prevention through continuous telemonitoring with real-time diagnostic value has yet to be achieved. Data on which machine learning (ML) models can be trained in order to detect or predict health conditions is sparse due to both mission and crew confidentiality and since astronaut crews are not currently equipped with 24/7 wearable medical monitoring technology. Additionally, since astronauts are generally healthy over short duration missions there is a lack of symptoms in their biomedical data. This work focuses on generating symptomatic wearable ECG data, specifically that of the Astroskin Biomonitor System, a wearable device, which has been used on several Human Research Exploration Analog (HERA) analog astronaut missions and tested on the International Space Station [2]. We present a deep learning framework which harnesses traditional ML to build a state-of-the-art generative model for astronaut health signals.

METHODOLOGY: Our framework is composed of encoder-decoder components for learning different representations of ECG signals that we fuse to generate new data, as well as classes, of ECG signals. The framework fuses both symptomatic morphology and domain specific characteristics - transferring ideas of style and content from the image processing domain to the time series do- main - to develop an effective novel feature extraction mechanism. After pre-processing, the discriminators are first trained on a signal to determine the “style” or “con- tent” of the signal. Thenceforth, the encoder-decoders are trained with their respective discriminators, not only to reconstruct the original signal, but also to maximize the confidence of the discriminator. For example, a “style” encoder-decoder-discriminator should be able to some- what reconstruct wearable asymptomatic signals, and the discriminator should have high confidence that the recon- structed signal is of the Wearable domain. In order to achieve this effect both discriminators followed by both the autoencoders are trained before the generator. For all the autoencoders and discriminators we performed a grid search for the number, size and type of layers (convolutional, feed-forward), activation functions (sigmoid, tanh, ReLU, Leaky-ReLU) and learning rate [0.00001, 0.001]. The ADAM optimization algorithm was chosen as it performed best during model training. Dropout layers were used and the model training grid was set between 100 and 500 epochs with the final model set to 200 epochs. The architecture that produced the best results for each encoder was: Conv1d(32)-LeakyReLU- Conv1d(16)-LeakyReLU-Dropout(0.4)-Conv1d(8)-tanh and decoder: Conv2DTranpose(32)-ReLU-Cov2D(32)- Relu-Conv2D(32)-Dense(256). The difference be- tween the encoder and discriminators was a change on the last layer and adding dense layers: Conv1d(32)-LeakyReLU-Conv1d(16)-Dropout(0.2)-Dense(1024)- Dense(2)-softmax.

RESULTS AND DISCUSSION: Our approach based on both reconstruction and the discriminators performs better than the baseline, which is informed only by the Mean Square Error (MSE). We strongly believe that what makes this method particularly effective is implementing an architecture of both style and content as opposed to being restricted to feature extraction algorithms. This approach outperforms the baseline by reproducing signals that are 3.6X closer to the original signals for wearable asymptomatic ECG signals, and 1.3X closer to the original signals for clinical symptomatic ECG data by producing more accurate samples of unseen classes, with about a 1.6X improvement in MSE. Our immediate next steps are to provide a comprehensive evaluation framework and metrics as well as compare the validation results by both an ML system and human specialised cardiologists.

[*] Equal Contributions

ACKNOWLEDGEMENTS: The authors would like to thank NASA, CSA and the SETI Institute for their continuous support during this work. The authors are also grateful for the computational resources provided by IBM.

REFERENCES:

[1] D. R. Williams, “The Biomedical Challenges of Space Flight,” *Annual Review of Medicine*, 2003.

[2] W. T. P. C. J. Villa-Col ́ın, T. Shaw, “Evaluation of astroskin bio-monitor during high intensity physical activities,” *Memorias del Congreso Nacional de In- genierıa Biomedica*, vol. 5, no. 1, pp. 262–265, 2018.

Requirements for Artificial Intelligence in clinical decision support system in space missions (from TCL4 to TCL5)

PI: Tianna Shaw, Exploration Medical Capability Element
tianna.l.shaw@nasa.gov
Team: Eleni Antoniadou, David Belo, Krittika D'Silva, Brian Russell, Frank Soboczenski, Brian Wang
NASA Frontier Development Lab

Artificial Intelligence (AI) will augment human performance in space & enable a safer clinical decision support system (CDSS)

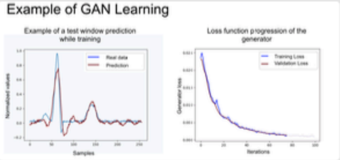
- As emerging technologies continue to gain traction and shift the way health assessment is performed, the space medicine industry has a tremendous opportunity to harness recent developments in AI to augment human performance.
- Our goal is to address the requirements so that astronauts can receive personalized biomonitoring and tailored directives to have their needs addressed in real-time as events unfold in space. This could include changes in radiation and microgravity levels during space travel.

Critical mission requirements to address with Artificial Intelligence-enabled CDSS

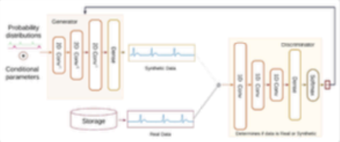
- AI has the potential to model multimodal, complex data (genomic, biomarker, phenotype, vital signs) to optimize crew performance and health
- Data that signify astronaut pathologies due to microgravity and radiation in space don't exist to train an AI diagnostic doctor
- Natural language processing can extract meaning from large datasets
- Different systems need significant time to train and much of the training is computationally expensive
- Minimize the need of network connectivity in extreme environments

Proof-of-concept study: AI enables the cardiac real-time telemonitoring and diagnosis from wearables in space

Synthetic data output from a generative model

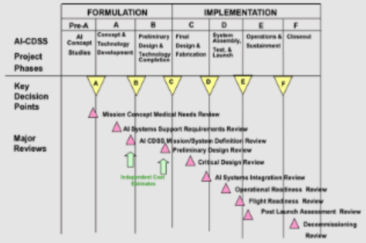


AI model architecture for ECG data synthesis from wearables

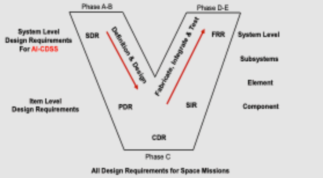


Architecture step-by-step requirements

Can we get AI-ready to increase effectiveness of clinical decision support systems for Lunar Missions and beyond?



Requirements Engineering Vee Diagram For Artificial Intelligence-CDSS



AI's role in clinical decision support

- **Two types of CDSS:**
  - Work with Knowledgebase
  - Work with Non-Knowledgebase
- **Knowledge based CDSS:**
  - Use knowledge from sources such as textbooks, and other resources
  - They have rules similar to if-then statements
- **Components of a knowledge based CDSS:**
  - Knowledgebase: Some source where they get their knowledge Inference engine: takes data and applies the rules from the knowledgebase
  - Communication: Allows system to communicate with user and user input
- **Many different AI techniques:**
  - Neural Networks are used to find patterns
  - Genetic Algorithms are used to find an optimal solution or diagnosis
  - Case Based Reasoning is used for evaluating data and learning from it
  - Natural language processing can be used to interpret large data corpuses

Conclusion

AI-CDSS systems enable:

- Autonomy in clinical prevention in space
- Real-time biomonitoring of an infinite amount of clinical data
- Improvement in patient outcome
- Higher Patient satisfaction
- Guidance for inexperienced practitioners
- Guidance interface for astronauts in deep space

AI-CDSS systems cannot:

- Replace a doctor/care giver
- Be 100% accurate/fool proof

APPENDIX E. ENVISIONING THE FUTURE ROLE OF AN EXPLORATION CLINICAL DECISION SUPPORT SYSTEM

This conference paper was written first authored by Bettina Beard with contributions from the research team at NASA, of which I am a member, on clinical decision support for deep space missions.

Authors : Bettina Beard, Brian Russell, William Toscano, Barbara Burian, Michael Krihak, Sandeep Shetye, Tiana Shaw

Conference Paper: Applied Human Factor and Ergonomics

Track : Neuroergonomic and Cognitive Engineering

Abstract : The Exploration Medical Capability Element of the NASA Human Research Program seeks to fuse new and existing technologies with practical mission goals into feasible and fiscally realizable human missions to the Moon and Mars. Expected communication delays with Earth-based medical experts will require unprecedented crew self-reliance to rapidly identify and treat anticipated and unforeseen medical conditions using constrained onboard resources with limited crew clinical skill. To extend the crew's capability to problem-solve and make decisions in stressful situations, all while performing other essential tasks, will require a three-pronged crew-system teaming approach. First, adaptive automation (AA), when implemented appropriately, can be an effective solution to workforce reduction (in this context, shifting from Mission Control expertise to a small crew with limited clinical skills) providing flexible support only when desired or required and based upon neurological indices and the situational context. Second, neurocognitive data can provide information about crew situation awareness, workload, fatigue, circadian desynchronies, while molecular genetics may bring insight into individual differences in spaceflight stressor effects (e.g., radiation exposure, elevated CO2, etc.). Third, human-system integration (HSI) performance-enhancing techniques including clinical

decision support systems (CDSS) with machine learning to assist the process of medical event classification, prediction, and pattern recognition.

The current project addresses the following research questions: (1) What is the current status of CDSS?, (2) What is the future role of an exploration CDSS?, 3) What aspects of existing CDSS can be leveraged to assist crew during exploration class missions?, 4) Can current knowledge about human cognition and neurocognition be used to enhance the efficacy of an exploration CDSS?, (5) How should NASA augment existing systems to address spaceflight-specific medical issues? 6) What are the specific data needs and information requirements of crew who have not received formal medical training?, and 7) What research, simulation and development is required to provide a collaborative, effective and reliable CDSS system for planned space exploration missions? During this session, the current status of this project will be presented. We will discuss how current modes of decision support (e.g., alerts of critical values, reminders of overdue preventive health tasks, guided clinical workflows, advice for drug prescribing, critiques of existing health care orders, and suggestions for various active care issues) can be tailored to exploration crew needs

APPENDIX F. SOFTWARE DEVELOPMENT OF ASSESSMENT TOOLS

Assessment Application on iPhone – COGNI

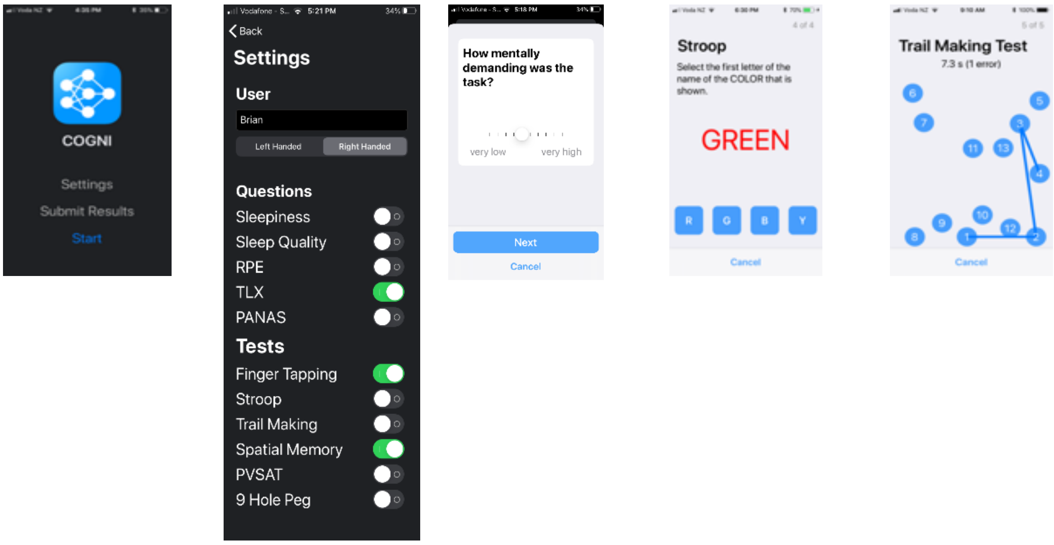
The iPhone app, named COGNI, was developed for the Apple iPad. Apple Research kit has built in set of cognitive assessments that can be modified into a protocol and has the capability of adding any custom questionnaire with Likert scales. The app was designed by myself and coded by a separate developer.

The application was designed for a single subject so that name and demographic data was entered only once at the start of the protocol. The duration of each assessment was configurable as well as which test and assessment to be used in a particular protocol.

The user work flow was designed to start the app and go directly into the protocol such that time and number of screen clicks was minimised. Font sizes where maximised for easy field reading and the layout was kept as simple as possible to be operated by fatigued participants.

The results were stored locally as the field protocol may not be cellular coverage and the results could be submitted to the researcher via email as a CSV file.

The start screen gives three options, settings, submit results and start test.



Settings

User

Brian

Left Handed | Right Handed

Questions

- Sleepiness ☐
- Sleep Quality ☐
- RPE ☐
- TLX ☑
- PANAS ☐

Tests

- Finger Tapping ☑
- Stroop ☐
- Trail Making ☐
- Spatial Memory ☑
- PVSAT ☐
- 9 Hole Peg ☐

How mentally demanding was the task?

very low — very high

Stroop

4 of 4

Select the first letter of the name of the COLOR that is shown.

GREEN

Trail Making Test

7.3 s (1 error)

5 of 5

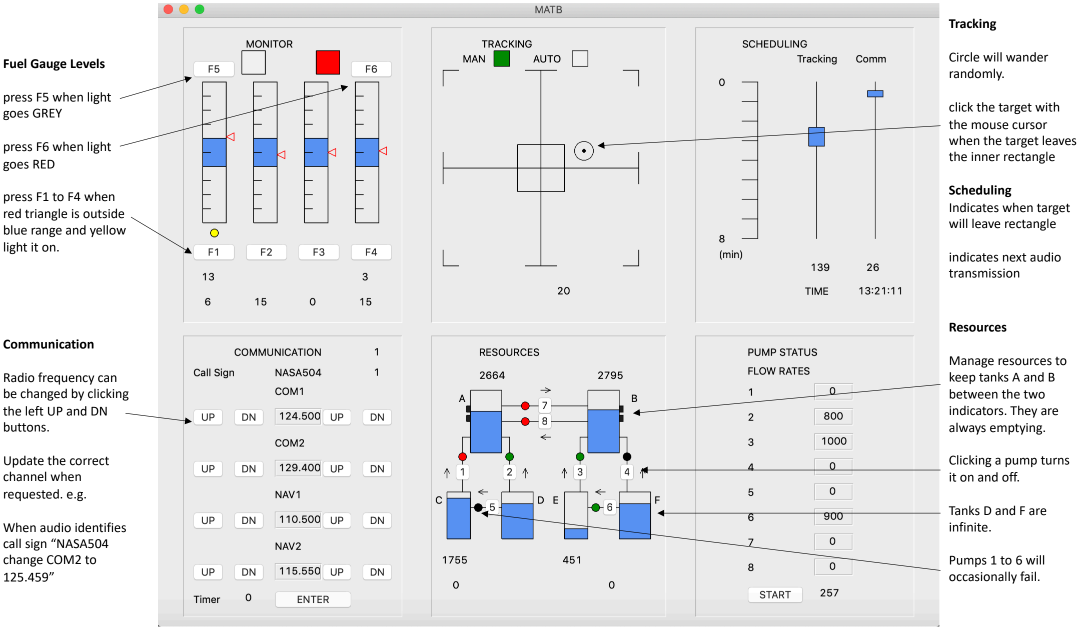
Home screen Settings Likert Test Stroop Trailing Making A

Figure A-1 COGNI App example screens

Cognitive Load Tool – Multi Attribute Task Battery, MATB

The MATB was developed at NASA by Comstock and Arnegard[123] to evaluate operator performance and workload. The program was rewritten in Python and PyQt by myself as part of this work based on the original paper. MATB has been used during sleep deprivation trials on fighter pilots[43,179].

The MATB app can be configured to run for a set amount of time when used as a cognitive load. The time between events can be defined in order to set a certain level of cognitive load. These were set experimentally in an unfatigued state in order to generate some errors to ensure sensitivity as the subject became more fatigued. Various responses to the different loads are recorded.



Fuel Gauge Levels

press F5 when light goes GREY

press F6 when light goes RED

press F1 to F4 when red triangle is outside blue range and yellow light it on.

Communication

Radio frequency can be changed by clicking the left UP and DN buttons.

Update the correct channel when requested. e.g.

When audio identifies call sign "NASA504 change COM2 to 125.459"

Tracking

Circle will wander randomly.

click the target with the mouse cursor when the target leaves the inner rectangle

Scheduling
Indicates when target will leave rectangle

indicates next audio transmission

Resources

Manage resources to keep tanks A and B between the two indicators. They are always emptying.

Clicking a pump turns it on and off.

Tanks D and F are infinite.

Pumps 1 to 6 will occasionally fail.

Figure A-2 Multi Attribute Test Battery