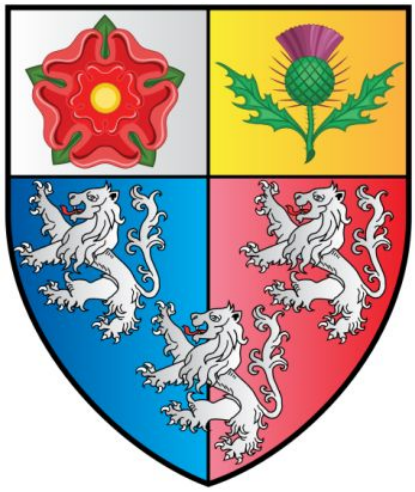
Inertial Sensors for the Assessment and Treatment of Shoulder Pain





Luckshman Bavan

Pembroke College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2020

Abstract

Subacromial shoulder pain is the third most common musculoskeletal complaint presenting to primary care. Despite the extensive amount of research carried out on the role of exercise-based rehabilitation, there remains uncertainty around the effectiveness of specific treatment strategies. Wearable sensors that enable practical monitoring of patient activity and objective measurement of movement could provide opportunities to address this problem. Magneto-inertial measurement units (MIMUs) offer clear advantages over existing motion capture systems, and can facilitate careful measurement of treatment implementation and response. There remains, however, a distinct lack of both technical and usability validation for the clinical application of these tools.

The aim of this DPhil thesis was to develop a practical MIMU based motion capture system and investigate its value in providing meaningful and useful feedback on shoulder movement and shoulder rehabilitation. This began with the development and testing of a processing pathway to derive reliable kinematic data from a commercial MIMU sensor under controlled motion conditions representative of human movement. The sensors were subsequently incorporated into custom-made wearable accessories, and shoulder motion data was collected from 70 healthy volunteers and 45 patients with subacromial shoulder pain. Processing of raw inertial data to clinically meaningful output was performed in MATLAB using a combination of custom-written and open-access code. Investigations were performed to evaluate the validity of a range of MIMU derived metrics for the assessment of shoulder activity and functional performance. Finally, the use of a prototype end-to-end remote monitoring system, offering near-real time feedback, was studied in 15 patients undergoing a course of home rehabilitation. System accuracy, usability and perceived impact of the technology was evaluated.

Metrics derived from an upper arm MIMU-based kinematic data demonstrated good psychometric properties in the assessment of patients with subacromial shoulder pain. Movement smoothness demonstrated reliability, known group validity and responsiveness to treatment. Metrics based on ranges of accelerometer and gyroscope data showed similar psychometric properties, as well as convergent validity, demonstrated by significant correlation with the Oxford Shoulder Score. Significant effect sizes for changes of these metrics following treatment were seen over a 12-week period. Implementing a single sensor setup, a fully automated processing pathway was validated for independent recognition and classification of shoulder exercise activity, with up to 97.2% overall accuracy. Quantitative and qualitative assessment of usability demonstrated that patients were comfortable with this form of monitoring and many felt these systems could improve their engagement with treatment.

MIMU based biofeedback systems offer a practical and reliable approach to measuring treatment compliance and recovery of function in patients with subacromial shoulder pain. They also have the potential to address factors that influence adherence, through self-monitoring and the provision of automated feedback. Through the development of novel data processing techniques and testing under clinically relevant conditions, this thesis has made several contributions to this field of study and provides translatable findings for researchers and clinicians aiming to employ wearable technologies for the assessment and treatment of musculoskeletal disease.

Acknowledgements

The content of this thesis is the result of three years of research at the Botnar Research Centre, Nuffield Orthopaedic Centre and Oxford Gait Laboratory. Many people have contributed to this work and I would like to express my sincere gratitude to them all. To begin with, I would like to thank my supervisors, Jonathan Rees, David Beard and Karl Surmacz. I am extremely grateful for all their guidance and support. Their collective knowledge and expertise provided me with a perfectly balanced platform to explore and contribute to this new and exciting field of science.

In addition to my supervisors, several other scientists and clinicians have inspired, guided or facilitated significant components of this work. Stephen Mellon has been very helpful throughout my research, and provided me with many opportunities to present, defend and refine my ideas. I am very grateful to Julie Stebbins and Andy Lewis for providing me with access and the training to use the Oxford Gait Laboratory. Cushla Cooper provided essential support with research ethics applications and the setup of clinical studies, and patient recruitment would not have been possible without the assistance of Colette Tracey, Bridget Watkins and Kim Wheway. I am also grateful to Michael Griffiths and the software team at McLaren Applied Technologies, who provided valuable support in developing the system used in my final study.

Finally, I would like to thank my wife, Sonia. This endeavour would simply not have been possible without her love and support.

Table of Contents

List of Abbreviations 8

List of Figures and Tables 9

1 Introduction 12

1.1 Problem identification 13

1.2 Thesis aim and objectives 15

2 Background and Literature Review 16

2.1 Aetiology and management of subacromial shoulder pain 17

2.1.1 Anatomy and function of the shoulder girdle 17

2.1.2 Aetiology of subacromial pain 19

2.1.3 Rehabilitation for subacromial shoulder pain 20

2.2 Wearable sensor-based assessment of the shoulder 22

2.2.1 MIMU sensor modalities 22

2.2.2 Wearable sensors for kinematic and functional assessment of the shoulder 24

2.2.3 Wearable sensors for rehabilitation activity classification 27

2.2.4 Wearable sensors for rehabilitation monitoring and feedback 30

3 Characterising Inertial Sensor Accuracy for Clinical Motion Analysis 32

3.1 Background 33

3.1.1 Sensor fusion for orientation estimation 33

3.1.2 Representing orientation 35

3.1.3 Determining the accuracy of MIMU systems 38

3.2 Study aims 40

3.3 Methods 41

3.3.1 Data Pre-processing 41

3.3.2 Experimental setup 44

3.3.3 Study protocol 46

3.4 Results 47

3.5 Discussion 52

3.5.1 Implications of study 52

3.5.2 Study limitations 56

3.5.3 Summary 57

4 Assessment of Shoulder Range of Motion 58
4.1 Background 59
4.1.1 Considerations for 3D modelling of the shoulder 59
4.2 Study aims 62
4.3 Methods 63
4.3.1 Movement range calculation 63
4.3.2 Study participants and setup 64
4.3.3 Study protocol and statistical analysis 65
4.4 Results 66
4.5 Discussion 69
4.5.1 Implications of study 69
4.5.2 Study limitations 70
4.5.3 Summary 71
5 Instrumented Assessment of Shoulder Function 72
5.1 Introduction 73
5.1.1 Functional assessment in the clinical setting 73
5.1.2 Measuring clinical function with inertial sensors 74
5.1.3 Validation of clinical assessment tools 75
5.2 Study aims 76
5.3 Methods 77
5.3.1 Performance scores 77
5.3.2 Study participants 80
5.3.3 Assessment protocol 81
5.3.4 Statistical analysis 82
5.4 Results 84
5.4.1 Reliability 86
5.4.2 Known-group validity 86
5.4.3 Score responsiveness 88
5.4.4 Concurrent validity 89
5.5 Discussion 91
5.5.1 Implications of study 91
5.5.2 Study limitations 94
5.5.3 Summary 94
6 Rehabilitation Exercise Monitoring and Classification 96
6.1 Introduction 97
6.1.1 Adherence in musculoskeletal rehabilitation 97
6.1.2 Supervised machine learning for human activity recognition 98

6.2 Study aims ... 99
6.3 Methods ... 100
6.3.1 Study participants and data collection ... 100
6.3.2 Pre-processing ... 102
6.3.3 Segmentation ... 102
6.3.4 Feature extraction ... 105
6.3.5 Feature selection ... 105
6.3.6 Model training and validation ... 107
6.4 Results ... 114
6.4.1 Reliability of segmentation ... 114
6.4.2 Feature selection ... 115
6.4.3 Model training validation ... 117
6.5 Discussion ... 120
6.5.1 Implications of study ... 120
6.5.2 Study limitations ... 122
6.5.3 Summary ... 123
7 Evaluation of Sensor-Supported Rehabilitation ... 124
7.1 Introduction ... 125
7.1.1 Connected health solutions for musculoskeletal rehabilitation ... 125
7.1.2 A prototype system design ... 125
7.1.3 System output ... 128
7.2. Study aims ... 129
7.3. Methods ... 129
7.3.1 Processing code development ... 129
7.3.2 Study participants and setup ... 130
7.3.3 Evaluation of activity recognition ... 132
7.3.4 Performance scores ... 132
7.3.5 Usability and perceived impact ... 133
7.4 Results ... 134
7.4.1 Accuracy of activity recognition ... 134
7.4.2 Performance score profiles ... 136
7.4.3 Usability ... 138
7.4.4 Perceived impact ... 141
7.4.5 Improvements ... 143
7.5 Discussion ... 144
7.5.1 Implications of the study ... 144
7.5.2 Summary ... 146
8 Discussion ... 147
8.1 Summary of main findings and conclusions ... 148

8.2 Limitations 151
8.3 Recommendations for future work 153
Appendix A 157
A.1 Device specifications 158
A.2 Orientation error data 159
Appendix B 162
B.1 Ethical approval 163
B.2 Design of wearable accessory 168
Appendix C 169
C.1 Feature rankings 170
Appendix D 173
D.1 MATLAB code 174
D.2 Interview topic guide 189
References 191

List of Abbreviations

3D	Three-dimension
10CV	Ten-fold-cross-validation
ADL	Activities of daily living
AUC	Area under curve
BF	Body frame
BLE	Bluetooth low energy
BMI	Body mass index
CI	Confidence interval
DCM	Direct cosine matrix
DoF	Degrees of freedom
DT	Decision tree
HAR	Human activity recognition
ICC	Intraclass correlation coefficient
ISB	International Society for Biomechanics
k-NN	*k*-nearest neighbour
LOSOV	Leave-one-subject-out-validation
MCID	Minimally clinically important difference
MEMS	Micro-electromechanical systems
MIMU	Magneto-inertial measurement unit
OMCS	Optical motion capture system
OSS	Oxford Shoulder Score
PPV	Positive predictive value
PROM	Patient reported outcome measure
RCT	Randomised controlled trial
REC	Research ethics committee
RF	Reference frame/Random Forests
RMSE	Root mean square error
ROC	Receiver operating characteristic
RoM	Range of motion
SD	Standard deviation
SRE	Sensor-to-reference error
SSE	Sensor-to-sensor error
SST	Simple Shoulder Test
SUS	System Usability Scale
SVM	Support vector machine

List of Figures and Tables

Figure 2.1. Anatomy of the shoulder girdle. 18
Figure 2.2. Shoulder motion described along three orthogonal planes of movement. 19
Figure 2.3. Typical MIMU sensor with tri-axis accelerometer, gyroscope, magnetometer. This device has Bluetooth streaming capability and a rechargeable battery..... 23
Figure 2.4. Reliable patterns in sensor data can characterise specific functional movements. Classification algorithms can be used identify more subtle patterns, without any user input. 28
Figure 3.1. Orientation of object determined as body frame relative to reference frame. 36
Figure 3.2. A rigid body undergoing [90° 90° 90°] rotation via different sequences, resulting in different orientations. 37
Figure 3.3. Any 3D rotation can be achieved with a single step rotation around the Euler axis. 38
Figure 3.4. Overview of processing pipeline developed and tested within this study chapter and implemented in future work. 41
Figure 3.5. Raw and pre-processed accelerometer data after bidirectional low-pass filtering. Filtered data does not contain high-frequency noise and recorded events are not delayed relative to the original signal. 43
Figure 3.6. (a) Photograph of the oscillating platform from the front and a view of a sensor secured in the main housing. Two sensors could both be contained in the housing stacked on top of one another. 45
..... 45
Figure 3.6. (b) Schematic diagram of the rotating platform, demonstrating the rotation possible in two axes. Encoders within the platform measure rotation relative to the starting orientation in both axes. 45
Figure 3.7. (a) MIMU orientation estimates compared with digital encoder data during section of dynamic attitude trial; (b) Normal distribution of recorded error in MIMU output the over entire five-minute trial. 48
Figure 3.8. Bar charts displaying median SRE under different conditions in attitude and heading. Error bars display 95% upper limit. 49
Figure 3.9. Box plots (centre bar: median; hinges: 1st & 3rd quartile; whiskers: 1.5 x interquartile range) comparing error during single-axis and multi-axis 1Hz trials for (a) attitude and (b) heading movement..... 50
Figure 3.10. Box plots (centre bar: median; hinges: 1st & 3rd quartile; whiskers: 1.5 x interquartile range) comparing error during *Early* (T_1) and *Late* (T_2) portions of 1Hz trials for (a) attitude and (b) heading movement. 51
Figure 3.11. (a) Orientation trace of two MIMUs undergoing the same movement demonstrates time delay in MIMU1 output; (b) Resulting orientation error between both MIMU outputs. 54
Table 3.1. Comparison of median errors and 95% upper limits of Madgwick, Kalman and complementary filter. All dynamic trials involved 1.0Hz ±10° oscillatory movement. 56
Figure 4.1. With 3D positional information of two sensors aligned to humerus and thorax, humerothoracic joint angles can be reported in three planes of movement, e.g. *abduction, flexion* and *rotation*. 61

Figure 4.2. Relative movement of the upper arm segment can be calculated from sensor orientation data at two different positions. Alignment of coordinate systems are not required. 62
Figure 4.3. Peak analysis of data streams was performed using the dominant axis of the Euler angle trace. 64
Figure 4.4. Study set up with wearable MIMU and retroreflective markers. A torso sensor was also trialled in several subjects however the data from this sensor was not considered for this analysis. 65
Table 4.1. Summary of results comparing MIMU and OMCS calculations of relative shoulder RoM. 67
Figure 4.5. Single axis orientation of arm reported by MIMU sensor and OMCS over time for two subjects. MIMU rotation plane measurements are notably underestimated relative to abduction and flexion plane measurements. Drift in MIMU reference frame position demonstrated in abduction and rotation traces for Subject 2. 68
Table 5.1. A range of end-point kinematic parameters classified by the movement characteristics they represent. 75
Figure 5.1. Single-axis angular velocity profiles (left column) during shoulder elevation and the corresponding Fourier magnitude spectra (right column). The bottom row profile displays reduced smoothness with greater number of submovements. The increase in submovements is reflected in the complexity of the frequency spectrum. 78
Figure 5.2. Tri-axis angular velocity profile (left) during shoulder elevation and plot of corresponding normalised vector (right). 79
Figure 5.3. Inertial sensor setup with (a) start and target end positions for (b) back and (c) bulb movements. Procedure performed at a self-selected pace and repeated three times for both arms. 82
Table 5.2. Demographic data and summary scores for matched control and patient groups at baseline and follow-up. columns represent non-dominant or affected shoulders, columns represent dominant or unaffected shoulders for control subjects and patients, respectively. 85
Figure 5.4. Box plots (centre bar: median; hinges: 1st & 3rd quartile; whiskers: 1.5 x interquartile range) for all asymmetry scores comparing patients (baseline scores) to control subjects for both *back* and *bulb* procedures. 87
Figure 5.5. ROC curves with AUC value reported for individual shoulder scores and asymmetry scores discriminating between patients and healthy controls. 88
Figure 5.6. Box plots (centre bar: median; hinges: 1st & 3rd quartile; whiskers: 1.5 x interquartile range) for patient individual shoulder scores (affected side) comparing baseline to follow-up. 89
Figure 5.7. Scatter plots demonstrating the relationship between asymmetry scores and the Oxford Shoulder Score. 90
Figure 6.1. Overview of activity classification study implementing supervised machine learning techniques. 98
Figure 6.2. Ten repetitions of five exercises recorded with a single arm mounted MIMU. . 101
Figure 6.3. (a) 3-axis gyroscope plot; (b) Peak analysis of (inverted) gyroscope vector magnitude identifies locations close to 0; (c) Locations of identified peaks superimposed on orientation data indicates segmentation points. 104
Figure 6.4. Feature mapping of *k*-nearest neighbour algorithm with dashed lines representing distances calculated between new class data and neighbouring training data. Relative distances between neighbouring training data classes will determine class allocation for test data. 108
Figure 6.5. Simple decision tree. 109

Figure 6.6. Random forests algorithm with individual trees trained on data bootstrapped from original data set. 110
Figure 6.7. Support vector machine with hyperplane created within largest boundary found between two classes. 111
Figure 6.8. *k*-fold cross-validation, $k = 10$. 113
Figure 6.9. Leave-one-subject-out cross-validation. 113
Figure 6.10. (a) All features ranked according to ReliefF algorithm weights (high to low) with individual plots highlighting specific feature subsets. (b) Relationship between feature inclusion (in order of ranking) and classifier performance (10FCV). 116
Table 6.1. Performance of all four classification models using 10FCV and LOSOV. 118
Figure 6.11. Confusion matrix of (a) 10FCV and (b) LOSOV validation results of single model (Random Forests). 119
Figure 7.1. System architecture of the wearable motion capture platform. 126
Figure 7.2. User interface (a) when registering user, (b) initiating exercises, (c) during exercise recording, (d) for reviewing adherence data and (e) for reviewing performance metrics. 127
Figure 7.3. Subacromial pain rehabilitation guide published in 2018 by British Shoulder and Elbow Society with support from the NHS Get it Right First Time (GIRFT) Programme. . 131
Table 7.1. Classifier sensitivity, PPV and specificity for each activity class. 135
Figure 7.4. Confusion matrix for classifier testing results. 135
Figure 7.5. Bar charts for each exercise comparing exercise count from the wearable sensor to true number of repetitions performed. 136
Figure 7.6. Power, smoothness and RoM scores following all rehabilitation sessions for select (a) patients with reported improvements and (b) patients with no reported improvements in pain and function. 138
Table A1. Metamotion R technical specifications. MetaMotion R, MbientLab, San Francisco, California, USA. 158
Table A2. Attitude and heading axis median error and 95% upper limit. Sensor-to-reference (SRE) and sensor-to-sensor error (SSE) reported over 300min trials. 159
Table A3. Multi-axis median error and 95% upper limit. Sensor-to-reference and sensor-to-sensor error reported over 300min trials. 160
Table A4. *Early* (T_1) and *Late* (T_2) portions of 300 min 1Hz trials for attitude and heading axes. Median error and 95% upper limit reported for sensor-to-reference readings. 160
Table A5. Median error and 95% upper limit in attitude and heading planes during 60 min static trial. 161
Figure B.1. Sponsor letter from the Clinical Trials and Research Governance team. 163
Figure B.2. Health Research Authority letter of approval. 167
Figure B.3. MIMU sensor incorporated into a wearable accessory. 168
Table C1. Time and frequency domain features derived from accelerometer, gyroscope, magnetometer and orientation data segments. Features were calculated from all axes and normalised vector magnitude values of all sensor modalities. 170
Table C2. All features ranked from highest to lowest based on ReliefF algorithm weight. . 172

1 Introduction

Contents

1.1 Problem identification 13
1.2 Thesis aim and objectives 15

1.1 Problem identification

Shoulder related disorders affect up to 14% of the UK population and are a very common cause for presentation to primary care (1). Subacromial pain, accounting for 70% of shoulder problems, can significantly compromise upper limb function, affect sleep and interfere with psychological and social well-being (2). Management of subacromial shoulder pain places a substantial burden on healthcare services, with a mean annual cost of the condition estimated at €4139 per patient (3). Operative management in England over the 10-year period between 2007 and 2017 was reported to cost just under £1.1 billion (4). Significant economic costs are also incurred secondary to the condition's impact on employment absences and workers' compensation pay-outs (5, 6).

Surgical options for treatment of subacromial pain have recently come under question after publication of trials assessing the efficacy of arthroscopic decompression surgery; concluding that the procedure offers no benefit when compared to placebo surgery (7, 8). Exercise-based rehabilitation therefore remains the cornerstone of management, and the benefits of these treatments have been demonstrated (6). However, there is no consensus on what is the best rehabilitation strategy, and it is unclear what quantity, quality and form of exercise programme is necessary to achieve optimal outcomes (9, 10). As a result, there remains huge variation in the delivery of musculoskeletal physiotherapy services (11).

The requirements of treating a rapidly growing and aging population with increasing functional demands are also placing immense strain on physiotherapy services, and in many cases, compromising the availability and consistency of rehabilitation. A consequence is the

increasing need for patients to undergo rehabilitation independently and unsupervised (11). While consistent engagement in home-exercise programs is a strong predictor for treatment success, adherence is often very low. It is also found that patients are less likely to perform their exercises correctly when unsupervised (12, 13). Delivering effective rehabilitation and evaluating the effectiveness of any particular rehabilitation strategy can be challenging when non-adherence is as high as 70% (14). Therefore, increased efforts are required to develop methods to measure and optimise adherence, and investigate rehabilitation treatments in a controlled manner.

Recent advancements in wearable sensor technology are providing opportunities to efficiently measure and monitor shoulder joint kinematics, and a growing body of research is exploring the applicability of these instruments in clinical settings and rehabilitation (15). Sensors to monitor and analyse human motion have the potential to measure and track clinical progress, as well as provide patients with meaningful real-time feedback, enabling effective self-directed therapy (16). Magneto-inertial measurement units (MIMUs) incorporate different sensor modalities to provide information on the position and motion of the sensing device. The performance of these wearable sensors have dramatically improved over recent years resulting in their widespread clinical use in gait assessment, falls detection and activity recognition (17). Qualitative studies also indicate that clinicians involved in the management of musculoskeletal conditions, including physiotherapists and orthopaedic surgeons, feel wearable technologies could positively complement their role and enhance their interaction with patients (18-20).

1.2 Thesis aim and objectives

The aim of this DPhil thesis was to develop a MIMU based motion capture system for the shoulder and investigate its practicality in providing meaningful and useful feedback on shoulder movement and shoulder rehabilitation.

This aim was realised by achieving the following objectives:

Develop a processing pipeline to derive stable orientation data from a commercial wireless MIMU sensor, and characterise the measurement accuracy under controlled dynamic conditions.

Determine the accuracy and reliability of a practical MIMU based wearable system in measuring shoulder joint range of motion.

Examine the potential for MIMU derived kinematic data to measure clinical shoulder function in patients with subacromial pain.

Investigate supervised machine learning techniques for recognition and classification of shoulder rehabilitation exercise activity from MIMU data.

Develop a prototype wearable motion capture system for rehabilitation monitoring in the home environment, and evaluate its performance in terms of technical accuracy, usability and perceived value.

2 Background and Literature Review

Contents

2.1 Aetiology and management of subacromial shoulder pain 17
2.1.1 Anatomy and function of the shoulder girdle 17
2.1.2 Aetiology of subacromial pain 19
2.1.3 Rehabilitation for subacromial shoulder pain 20
2.2 Wearable sensor-based assessment of the shoulder 22
2.2.1 MIMU sensor modalities 22
2.2.2 Wearable sensors for kinematic and functional assessment of the shoulder 24
2.2.3 Wearable sensors for rehabilitation activity classification 27
2.2.4 Wearable sensors for rehabilitation monitoring and feedback 30

2.1 Aetiology and management of subacromial shoulder pain

2.1.1 Anatomy and function of the shoulder girdle

The shoulder girdle connects the upper limb to the axial skeleton and is composed of multiple joints. The complexity of its structure allows for a large range of movement in multiple planes. However, this mobility comes at the expense of reduced stability and a greater tendency for dislocation or injury (21). The joints that form the shoulder girdle include the glenohumeral joint, acromioclavicular joint, sternoclavicular joint and scapulothoracic joint. The sternoclavicular joint is a synovial saddle joint that connects the manubrium of the sternum to the clavicle. It is the only true joint to link the upper limb to the axial skeleton and offers very limited mobility. The acromioclavicular joint is a plane synovial joint connecting the clavicle to the acromion of the scapula and also offers limited mobility.

The scapulothoracic joint refers to the articulation between the scapula and the thorax. It is not a true joint, although correct and well-coordinated movement at this interface is critical to healthy functional movement of the upper limb. Finally, the glenohumeral joint is the articulation between the humeral head and the glenoid fossa of the scapula. This is a ball and socket joint and naturally offers the greatest degree of mobility within the shoulder girdle. The articulation is very shallow, allowing the fossa to accommodate for less than one third of the humeral head; therefore, the joint's stability is provided mostly by the surrounding soft tissue structures. This includes the labrum (a ring of fibrocartilage around the glenoid), glenohumeral ligaments, coraco-humeral ligaments and the rotator cuff muscles (22).

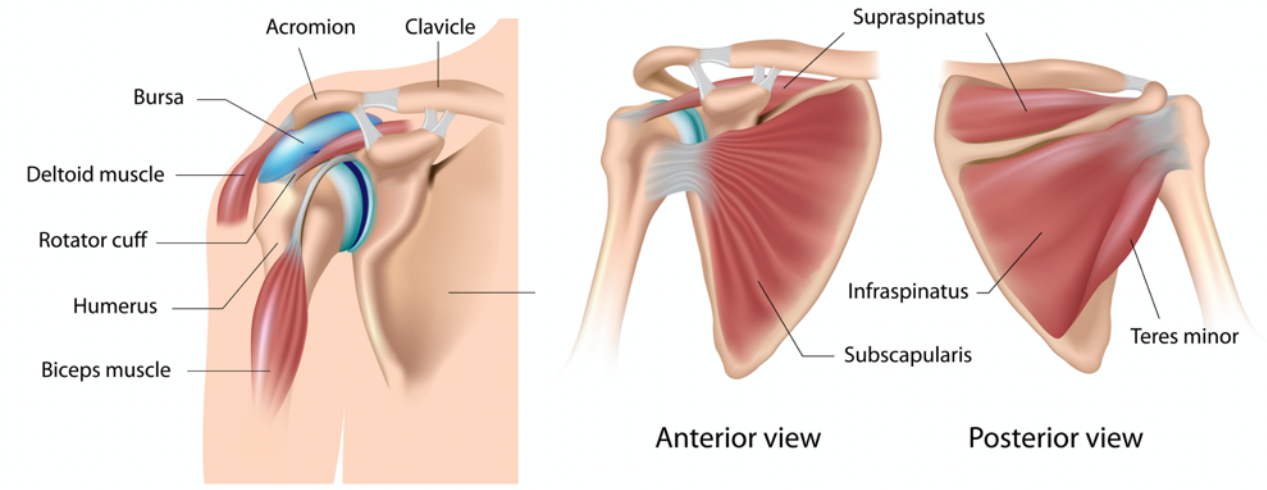


Figure 2.1. Anatomy of the shoulder girdle.

Movements of the shoulder joint include abduction/adduction (lateral and medial rotation of the upper arm along the sagittal plane), flexion/extension (forward and backward rotation of the upper arm along the coronal plane and internal/extension rotation (rotation of the humerus medially and laterally around its long axis). Most functional upper limb activities typically involve compound movements with simultaneously coordinated motion in multiple planes (23).

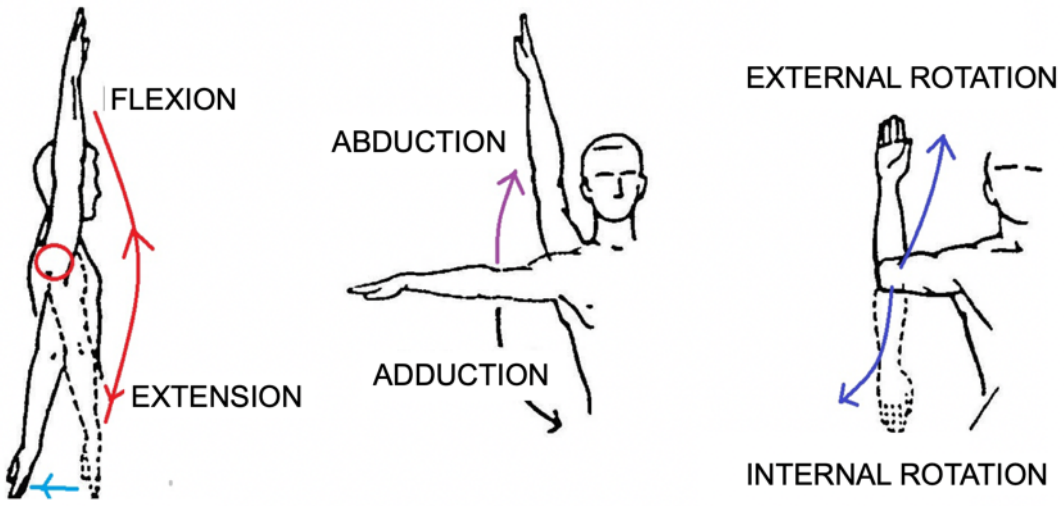


Figure 2.2. Shoulder motion described along three orthogonal planes of movement.

2.1.2 Aetiology of subacromial pain

Subacromial shoulder pain, sometimes referred to as shoulder impingement, is a degenerative condition involving injury to the structures within the subacromial space. The condition is commonly believed to be caused by irritation of the rotator cuff tendons due to mechanical narrowing of the subacromial space, or due to an increase in the volume of the subacromial soft tissues (24). In many cases, a bone spur can be identified at the anteroinferior margin of the acromion, which increases the chance of physical contact, particularly when the shoulder is abducted and internally rotated (25). To treat subacromial pain, Neer proposed decompression of the subacromial space; a procedure involving removal of the bone spur, which has often been combined with bursectomy and release of the coracoacromial ligament (26). Despite a lack of prospective evidence to support the use of subacromial decompression surgery, the number of patients undergoing the procedure in England increased seven-fold from 2000 to 2010 (27). Recent RCTs have demonstrated that this surgery does not offer any clinically important benefit over placebo surgery or no treatment at all, for the majority of patients undergoing decompression (7, 8).

The mechanical impingement explanation for subacromial pain is under dispute, as it cannot fully explain the aetiology or natural history of the condition (28, 29). The failure of decompression surgery to reliably relieve symptoms also confirms that this may be too simplistic an understanding of this condition. An updated explanation, and one that has gained recent popularity, is that both intrinsic and extrinsic factors contribute to subacromial pain and rotator cuff pathology (30, 31). Intrinsic factors relate to the physiological and mechanical properties of the rotator cuff tendons, influenced by age, vascularity and genetic predisposition. Extrinsic factors include anatomical factors such as the size and shape of the acromion and thickening of coracoacromial ligament. External factors also include biomechanical influences, such as changes in posture and altered glenohumeral and scapulothoracic kinematics, often caused by weakness and imbalance in the periscapular musculature (30, 32).

2.1.3 Rehabilitation for subacromial shoulder pain

Exercise-based rehabilitation aims to address some of these extrinsic factors through programs designed to improve muscle imbalance and postural dysfunction. Programs consist of scapular stabilisation, rotator cuff strengthening, range of motion and stretching exercises, and are typically initiated as first line treatment. Several reviews evaluating the effectiveness of exercise therapy for the treatment of subacromial pain have been published (6, 33-38). Some studies have demonstrated that exercise therapy can be effective in improving pain, function and range of motion, and three reviews have reported exercise therapy to be more effective than control or placebo interventions (35, 36, 38). The reviews do, however, generally report that the quality of available evidence is poor. In addition to this, a recent Cochrane review only identified a single good quality study comparing exercise treatment reflective of common

current practice to a placebo treatment (inactive ultrasound therapy); no clinically important differences were found between the outcomes of both groups (39). There is also uncertainty around what constitutes the optimal rehabilitation strategy, with multiple competing approaches being prescribed to patients. Very few trials have been conducted to compare different combinations of exerices and different forms of delivery, and reviews have concluded that there is insufficient evidence to support or refute the effectiveness of one strategy over another (9, 10).

Recommendations for enhancing the quality of rehabilitation research include providing clear descriptions of eligibility criteria, ensuring adequate sample sizes and appropriate lengths of follow-up. In addition to this, understanding the dose-response relationship has also been highlighted as a priority (10). To achieve this, studies require robust and reliable methods to measure treatment fidelity, which includes quantifying compliance and establishing treatment differentiation between alternative interventions. Measures of adherence are commonly lacking from clinical trials of rehabilitation for musculoskeletal conditions (13). When rehabilitation compliance is monitored, it is often achieved using crude measures such as attendence at appointments (12). These measures would not account for home exercise activity or even indicate if individuals are completing the expected tasks. The majority of intruments available for home compliance monitoring are questionnaires or log books based on patient self-reporting. Most of these approaches are of questionable validity or have not undergone any form of psychometric testing (12, 40), and several studies show that through self-reporting patients have a tendency to significantly overestimate home exercise adherence (41-43). These concerns relate not only to the quantity of rehabilitation adherence, but also to the quality of movement during exercise, which is an important component of treatment fidelity.

2.2 Wearable sensor-based assessment of the shoulder

2.2.1 MIMU sensor modalities

MIMUs incorporate micro-electromechanical systems (MEMS) sensing technology to measure rotational speed and orientation relative to the earth's gravitational and magnetic field (44). The three MEMS sensing modalities used for this work are detailed below and full device specifications are available in Appendix A.

Accelerometers detect *proper* acceleration, which is different to *coordinate* acceleration, and effectively means that these devices can be used to detect the direction of gravity. Gravitational acceleration dominates the raw sensor output, which is expressed in m/s^2 or g where $g = 9.81 m/s^2$. Accelerometers are susceptible to high frequency noise, which can compromise the short-term value of their readings. A horizontal reference frame can be created from the detected direction of the gravity vector, which allows for orientation estimation in the horizontal plane (*pitch* and *roll*) only (45).

M**agnetometers** quantify the local magnetic field to provide an absolute position relative to magnetic north, which enables a sensor to reference its rotation around a vertical axis. However, the accuracy of these sensors relies on a consistent magnetic field, which is easily disrupted by the presence of ferromagnetic materials. Magnetic field strength is expressed in Gauss (*Gs*) or micro Tesla (μT) (46, 47).

Gyroscopes measure angular or rotational velocity in up to three axes, typically expressed as degrees or radians per second. Integration of gyroscopic output allows for orientation estimation, however, accuracy is compromised by gyroscopic bias secondary to electronic noise and temperature effects. Integration of these imperfect signals causes a progressive accumulation in error, often referred to as *drift* (48). Unlike accelerometers and magnetometers, gyroscopes do not have an external reference.

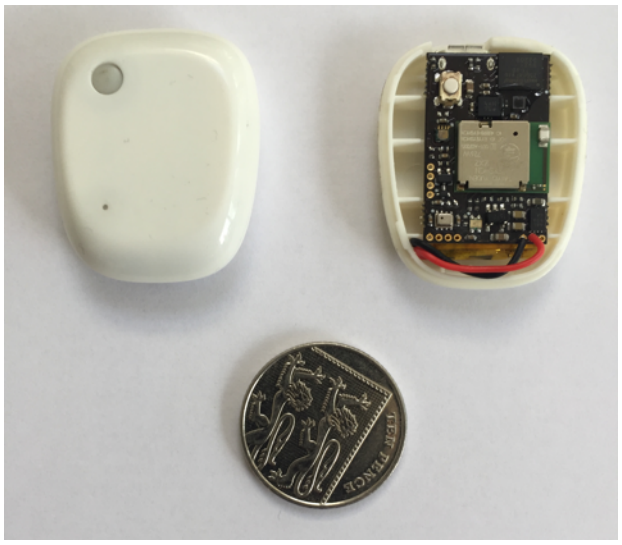


Figure 2.3. Typical MIMU sensor with tri-axis accelerometer, gyroscope, magnetometer. This device has Bluetooth streaming capability and a rechargeable battery.

The performance of wireless MEMs technology has improved over the past few decades, although referencing position is less accurate than using other technologies such as solid state accelerometers or optical gyroscopes (49). As described above, error in inertial measurement can originate from biases and thermo-mechanical noise, and intelligent processing is required to mitigate this error. The methods to derive stable and reliable measurements will be detailed in the following chapter.

2.2.2 Wearable sensors for kinematic and functional assessment of the shoulder

Many forms of motion assessment exist and are capable of measuring upper-limb kinematics (50-54). The translation of these biomechanical methods to opportunities and improvements in clinical practice is well documented (55). Kinematic measurements include range of motion (RoM), execution time and movement speed, but can also provide information on movement *quality*. Movement properties such as coordination, control and fluency can potentially reflect clinically important domains of function (56). There has been an increasing trend towards investigating the role of wearable systems in providing this data. This has been encouraged by the developments in MEMS technology, which have led to comfortable and unobtrusive methods to measure human movement in virtually any setting.

Extensive investigations have been carried out evaluating MIMUs' ability to report human joint angles, and many commercial systems offering to support patient rehabilitation are based on joint angle monitoring (57). Compared to the lower limb, kinematic measurements of the upper limb are complicated by the higher degrees of freedom and a lack of distinct movement patterns. Identifiable signatures during the gait cycle, such as heel strike or terminal swing, provide detail for reference and calibration (58). As detailed above, the movements of the shoulder occur through multiple articulations, and it is challenging to model each individual segment and articulation with surface marker or wearable motion technology. As a result, most MIMU based measurement of shoulder RoM is of the humerothoracic joint (humerus relative to the torso) (59). MIMU measurements which have been compared to gold standard optical reference system and have demonstrated root mean square error (RMSE) values below 12° (60-62), with the exception of one study reporting RMSE values up to 26° (63). Reliability of recordings have been assessed with intraclass correlation coefficients (ICCs), demonstrating

values between 0.68 and 0.96 (64, 65). While it is difficult to compare results across different studies due to fundamental differences in study protocol, modelling parameters, calibration and processing steps, the general trend is that abduction/adduction and flexion/extension movements are recorded more accurately and reliably than internal/external rotation movement (59). This is likely to be related to soft tissue artefact resulting in movement, or lack of movement, of sensors relative to underlying bone, and is a well-recognised problem for surface marker-based motion analysis (66). MIMU platforms have been applied to measure scapulothoracic and glenohumeral joints independently (67), however, motion artefact is potentially an even greater issue and the impractical experimental setups do not appear to have any scope for application in routine clinical settings or home environments. Measurement of humerothoracic joint angles has been achieved with more clinically acceptable wearable setups, and still offer relevant information about functional movement ranges. With acceptable levels of reliability, there is strong potential for the use of these systems in performance tracking.

Other quantitative measures of shoulder movement quality can be used to reflect outcomes following treatment and changes in a patient's functional state, and many of these parameters can be measured with standard MIMU devices. Studies evaluating these alternative kinematic parameters have done so with measurements taken during defined functional movements or complete ADLs. In many cases, studies have focused on collecting data from healthy volunteers with a view to establishing normative data values in relation to certain kinematic variables. Coley et al. first introduced the concept of measuring shoulder function in patients with common shoulder pathologies using scores derived from inertial sensor data (68). Utilising functional upper limb movements derived from those described in the Simple Shoulder Test (SST) and generating scores based on acceleration and angular velocity ranges,

authors demonstrated that inertial metrics could reliably differentiate pathological from healthy shoulder movement. Other studies have progressed this work, demonstrating that these scores can have moderate correlation with validated PROMs and are capable of tracking patient progress following treatment (69, 70). The focus of these studies is largely on a metric derived from the product of the ranges of acceleration and angular velocity; being a power dependent quality, this metric is intended to reflect the control strategy in arm movement. Range of angular velocity alone has also been investigated and was shown to have good psychometric properties in the evaluation of shoulder function (68, 71). Naturally these metrics will have some correlation with task duration, although it is unclear what speed and acceleration-based measures can offer in addition to task duration, which is a well-established and commonly used metric for the assessment of lower-limb clinical function in musculoskeletal patients (72, 73).

An alternative approach to assessing movement quality is to quantify *smoothness* (56, 74). Movement smoothness is a feature related to the non-intermittency of movement and can be observed through accelerometer or velocity profiles. Smoothness is largely considered a property of well-developed motor behaviour and can be related to numerous factors including the nature of the task, familiarity with the task and neuromuscular impairment (75-77). Common methods to quantify smoothness using inertial data include quantifying velocity peaks, quantifying rate of change of movement acceleration and analysing the frequency composition within velocity or acceleration profiles (78). The latter of the three methods has gained some interest in recent years, demonstrating good properties as a functional measure in clinical studies (74, 79, 80). While it is understood that factors at level of the musculoskeletal system also contribute to this movement quality, little is known about smoothness profiles of patients with musculoskeletal pathology, as clinical investigations have been largely focussed on patient populations suffering with neurological conditions. Osteoarthritis of the knee joint

has been associated with reduced movement smoothness during gait (81), and it is possible that other degenerative conditions could have a similar effect on movement smoothness.

2.2.3 Wearable sensors for rehabilitation activity classification

Monitoring and classification of human activity is a popular application for wireless inertial sensors. In the context of healthcare research, this has been applied for the assessment of ADLs to provide a measure of patient performance, the detection of abnormal events such as falls and the classification of gait patterns in neurological patients (82, 83). The ability to conveniently and objectively measure rehabilitation exercise adherence would be useful to understand factors around patient engagement, as well as evaluate the success or failure of specific rehabilitation exercises in an informed way. Quantifying the rehabilitation dose-response effect, with regards to frequency, intensity, pace and rest, can also help develop evidenced based treatments that are time and cost effective for both patients and healthcare providers. It is possible to estimate limb poses and positions via kinematic data, and some studies have explored this specifically for recognising rehabilitation activity. A powerful method to achieve this is through the use of supervised machine learning techniques. The application of these techniques to data acquired from wearable inertial sensors is well established in the field of human activity recognition (HAR) (84). The key benefit of using artificial intelligence-based algorithms is that it is not necessary to explicitly program the rules required to differentiate one activity from another. The relevant parameters and thresholds used to classify activity are established independently by models capable of *learning* from existing training data. These machine learning models or algorithms, used to solve classification problems, are often referred to as *classifiers*.

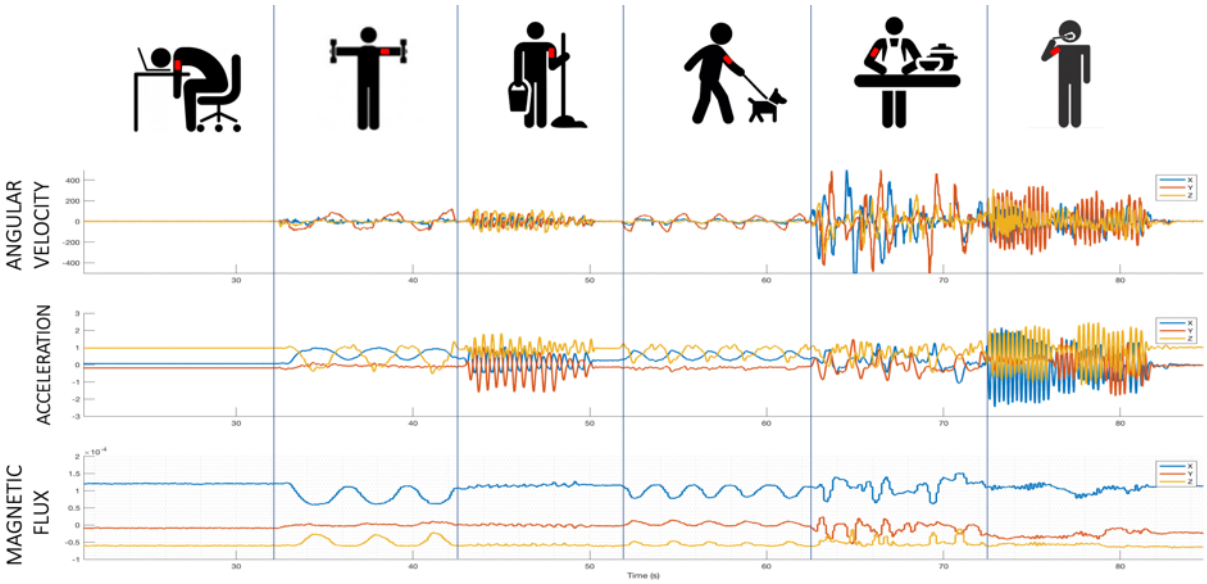


Figure 2.4. Reliable patterns in sensor data can characterise specific functional movements. Classification algorithms can be used identify more subtle patterns, without any user input.

As with general kinematic assessment, activity recognition applications for the upper limb face greater challenges when compared to the lower limb. The multiple degrees of freedom offered at the shoulder and forearm permit a far greater range of functional movements to be performed, and allow for different combinations of sub-movements to execute the same task. The higher degree of inter and intra-subject variability for ADL and rehabilitation activity performance naturally makes it difficult for algorithms to profile and classify movements based on kinematic data. The first studies evaluating inertial sensor technology combined with machine learning techniques to classify rehabilitation activity were focused on lower limb exercises (85, 86), and provided valuable insights into the general challenges associated with sensor placement, training data requirements and data processing. Pan et al. provided the first investigations into classifying shoulder rehabilitation exercises using body worn inertial sensor technology (87). Two tri-axis accelerometers were attached to the upper arm and torso, and a group of healthy volunteers and patients were monitored performing five shoulder exercises. Cross-validation testing demonstrated that a traditional machine learning algorithm could

accurately classify exercises at a rate of 92.6 to 99.1%, depending on the specific exercise. Authors did not report on the classification rates of healthy volunteers and patients separately, and the patients' shoulder conditions were not detailed.

Burns et al. performed a similar study reporting on classification rates of shoulder rehabilitation activity, performed by healthy volunteers (88). This study adopted a single sensor setup, utilising the tri-axis accelerometer and gyroscope within a commercially available smartwatch positioned at the wrist. Unlike the previous study, the authors evaluated a range of machine learning classifiers, reporting optimal accuracies from an ensemble classifier and a neural network algorithm. This study also investigated which sensor features were most useful for differentiating one activity from another, identifying accelerometer-based features to have greater predictive importance than features derived from the gyroscope. When models were tested on subjects who had previously contributed motion data used for training of those specific models, they demonstrated average predictive accuracies of up to 99%. When trained on data from all subjects except the test subject, the average predictive accuracy of the best model fell to 89%. These classification rates, achieved with a practical and patient friendly setup, would certainly suggest that remote shoulder exercise monitoring is technically feasible. Unfortunately, the potential for machine learning algorithms to classify exercise data in a real clinical setting cannot be fully realised until automated methods to filter out meaningless data and reliably segment exercise data are achieved. The classification pipeline used in this study, and previously mentioned studies, require varying degrees of manual intervention to segment data and present it to the machine learning classifier. Despite some studies implementing some form of automated segmentation of exercise data (85, 88), investigators have been required to manually remove non-specific data or screen and correct automated segmentation output. Additionally, no studies report the shoulder exercise classification rates for patients with

genuine pathology, who will display greater variation in movement than healthy volunteers. This raises further doubts regarding the application of this form of technology in real clinical scenarios, and highlights areas for further investigation.

2.2.4 Wearable sensors for rehabilitation monitoring and feedback

Connected health, also known as technology enabled care, refers to a model for healthcare delivery, based on the use of technology to provide remote healthcare services (89). The aim of this socio-technical model is to improve the use of healthcare resources and provide flexible opportunities for patients to engage with clinical services (90). Within rehabilitation, remote collection and aggregation of patient data has been suggested to provide benefits in terms of clinical outcomes and patient satisfaction, and can be used to encourage self-management of chronic conditions (16, 91). Sensing technology that is capable of detecting body movements can provide augmented biofeedback concerning quality or quantity of exercise execution, which can compliment and enhance supervised physiotherapy sessions, and also assist unsupervised rehabilitation in the home environment.

Various forms of technology supported rehabilitation have been investigated in musculoskeletal patients, where motion detection is achieved with infrared cameras, RGB cameras, pneumatic force sensors and inertial sensors (92-98). Study populations include patients following total knee replacement and anterior cruciate ligament reconstruction, and patients with lower back pain and frozen shoulder. Participation exercises have ranged from analytical exercises (e.g. joint movement not linked to any specific skills) to task-oriented exercises (e.g. developing skills and increasing levels of participation), and any real-time feedback provided to patients within these investigations were largely based on RoM during

exercises or tasks. Some studies have reported improved outcomes with technology supported rehabilitation, however the quality of the evidence is generally poor and studies are not adequately powered. Also few studies have implemented sensor biofeedback systems in an unsupervised environment. Ayode et al. published an investigation into the use of an inertial sensor-based platform for postoperative total knee replacement rehabilitation at home (92). The platform involved measuring movement during single plane knee exercises via the use of two wireless inertial sensors linked to a laptop. A custom made interactive software encouraged the performance of exercises and provided feedback on RoM and compliance. The authors reported that this system improved patient motivation and functional recovery.

Several barriers to delivering effective rehabilitation and evaluating its impact were highlghted earlier in this chapter. Connected health solutions based on wearable MIMU sensors offer great potential to overcome these barriers, due to their capacity to measure human movement in a practical and affordable way. This thesis intends to explore this potential for the benefit of assessing and managing patients with subacromial shoulder pain.

3 Characterising Inertial Sensor Accuracy for Clinical Motion Analysis

Contents

3.1 Background 33
3.1.1 Sensor fusion for orientation estimation 33
3.1.2 Representing orientation 35
3.1.3 Determining the accuracy of MIMU systems 38
3.2 Study aims 40
3.3 Methods 41
3.3.1 Data Pre-processing 41
3.3.2 Experimental setup 44
3.3.3 Study protocol 46
3.4 Results 47
3.5 Discussion 52
3.5.1 Implications of study 52
3.5.2 Study limitations 56
3.5.3 Summary 57

3.1 Background

Investigation of Magneto-inertial measurement unit (MIMU) capability for human kinematic assessment and activity recognition has become a prominent field of research, with a growing body of work focussing on validating MIMU-based systems at a clinical level (15, 59, 99). However, clinical settings are difficult to control, and error can be introduced from positioning, calibration, skin-motion artefact and individual movement variability (100). Systematic and controlled investigation of MIMU-based systems using mechanical or robotic platforms provide researchers with a baseline understanding of intrinsic sensor capabilities and the factors influencing accuracy (101, 102). Despite the obvious benefit this information will have in profiling system errors and understanding limitations at a clinical study level, far less attention is paid to this field of work. This chapter describes the pre-processing required to derive operational inertial data and calculate stable 3D orientation data from raw sensor signals (Figure 3.1). Methods to represent orientation and measure the rotation of a rigid body are also discussed. The study aimed to characterise the orientation accuracy of a commercial MIMU sensor implementing a series of custom-designed pre-processing steps. Testing was performed under a variety of controlled dynamic conditions to understand the potential strengths and weaknesses of MIMU assessment of human shoulder movement.

3.1.1 Sensor fusion for orientation estimation

Pre-processing can only achieve a certain degree of data correction, which would still not permit a direct calculation of a stable drift-free 3D orientation. It is also important to highlight that characteristics in gyroscope bias change over time. *Sensor fusion* offers the

opportunity to combine imperfect sensory data to produce an output that benefits from the strengths of each sensor modality. An *orientation filter* is a sensor fusion algorithm designed to output a stable and accurate orientation calculated from noisy and biased accelerometer, gyroscope and magnetometer data.

Various algorithms differ in how gyroscopic drift, inertial acceleration and magnetic disturbances are accounted for, and how orientation is described. *Kalman filters* are based on a set of recursive equations used to gain an estimate of a system state, despite the presence of significant error in real-time measurements. The process consists of 1) making a prediction of orientation based on previous measurements, 2) comparing the prediction with current sensor measurements and 3) updating the system's understanding of the relationship between prediction and recorded position. The discrepancy between the predicted and measured state is moderated based on an estimation of error in the system and the measurements, and the cycle is repeated. The Kalman filter has become the accepted basis for the majority of MIMU orientation filters (103, 104) and many commercial systems are centred on its use. However, these fusion algorithms are computationally expensive, require tuning of multiple parameters and typically demand high sampling rates (105).

An alternative category of orientation filter, that is computationally less expensive, is the *complementary filter* (106). As described previously, the accelerometer and magnetometer signals are susceptible to short term disturbances. However, over relatively longer periods of time, they do provide an accurate averaged representation of *attitude* (horizontal plane orientation) and *heading* (vertical plane orientation), respectively. In contrast, integration of gyroscopic readings cannot be trusted over longer periods due to the potential for drift, but gyroscope output is very accurate in the short term for representing angular rate. In simple

terms, the complementary filter combines estimates from low-frequency accelerometer output and high-frequency gyroscope output to derive stable orientation. In 2010, Madgwick et al. introduced an alternative complementary filtering approach (103). The error between both the gravity and magnetic field vector and the integrated gyroscope signal is minimised in an iterative *gradient descent* algorithm. The filter incorporates magnetic distortion and gyroscopic drift compensation, and only requires tuning of two parameters. Similar to alternative complementary filtering techniques, the Madgwick filter is computationally inexpensive, which strongly supports its use for real-time clinical applications (103).

3.1.2 Representing orientation

Describing the orientation of an object is equivalent to describing its rotation relative to a reference coordinate system (107). The figure below shows a coordinate system, which describes the reference frame (RF). In order to describe the orientation of the object within the figure relative to the RF, it must also have a fixed coordinate system, and this can be referred to as the body frame (BF). Determining and describing the relative 3D orientation of two frames of reference is the underlying task for most real-world orientation problems, including those within navigation, robotics and human biomechanical modelling. The orientation of BF relative to RF can be described in a three by three matrix known as a *direct cosine matrix* (DCM), which is calculated based on the geometrical relationship between the two frames. *Parameterisation* of DCMs is often performed to express orientation in alternative ways, the most common forms being *Euler angles* and *quaternions*. Parameterisation allows orientation to be represented in fewer than nine elements, reducing the complexity of any calculations, and allowing for a more interpretable description of rotation and orientation.

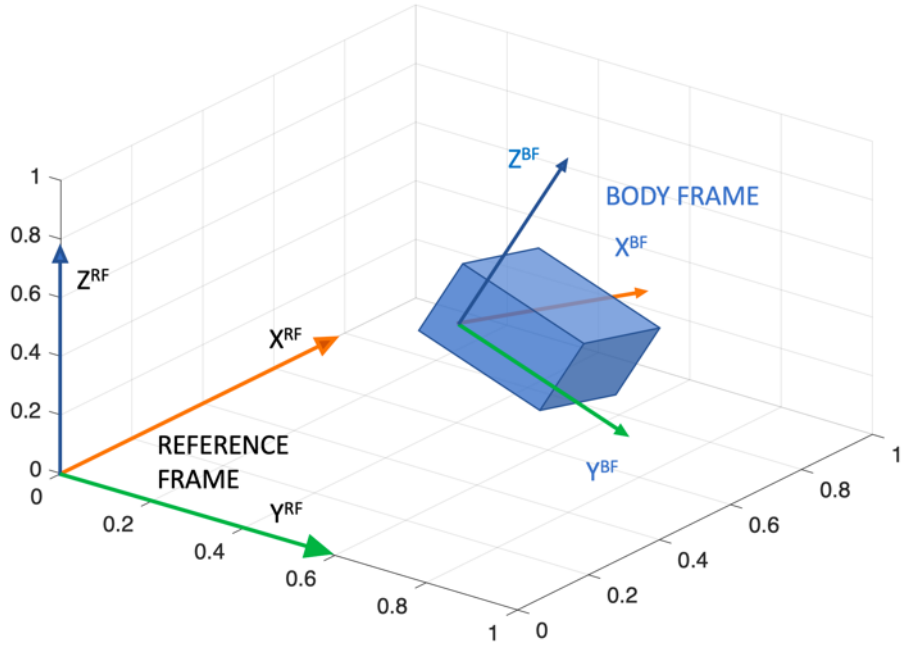


Figure 3.1. Orientation of object determined as body frame relative to reference frame.

Euler angles are the easiest to interpret as they provide an intuitive sense of orientation, with rotation reported as three orthogonal angles, often expressed as *pitch*, *roll* and *yaw*, or *x*, *y* and *z*. Pitch and roll typically describe rotation in two horizontal axes (attitude), and yaw describes rotation around a vertical axis (heading). Euler angles essentially refer to three rotation angles one can use to describe the rotation of a rigid body in space or describe a rigid body's position relative to a specific coordinate frame (e.g. RF). It is possible to apply the three rotations in a total of 12 different sequences around two or three orthogonal axes (e.g. ZYX, ZYZ, YXZ) to achieve a specific orientation (107). Therefore, it is essential to specify the sequence of rotation that is used. A key deficiency with the use of Euler angles is their inability to describe certain orientations. This can be best understood if we consider Euler angles representing three axes on a gimballed system. A problem known as *gimbal lock* occurs when two axes are driven into a parallel configuration rendering one of the three axes ineffective. Only two specific configurations result in exact gimbal lock; however, for most systems, problems are experienced as axis orientations begin to approach these configurations.

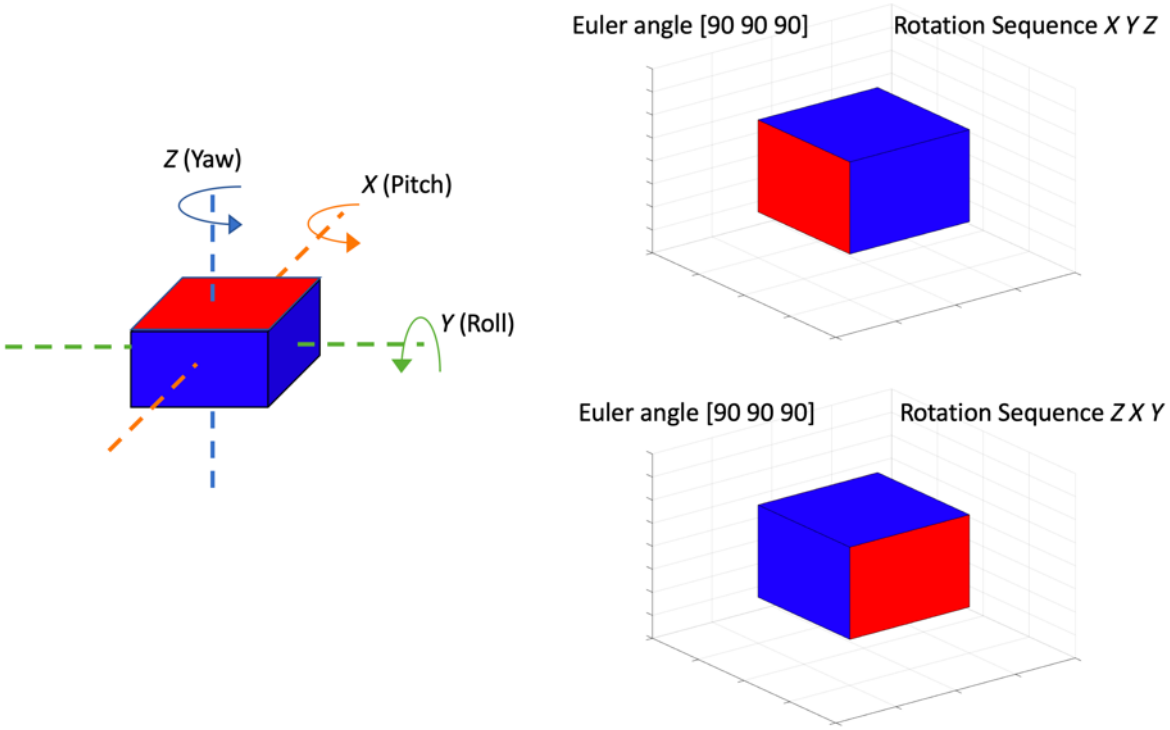


Figure 3.2. A rigid body undergoing [90° 90° 90°] rotation via different sequences, resulting in different orientations.

Quaternions consist of four parameters rather than three and are typically expressed in the form: $q = [w\ x\ y\ z]$, where w is a real number and x, y, and z are composed of both real numbers and complex numbers. Any sequence of rotations of a rigid body can be achieved with a single rotation by a given angle θ around a fixed axis (*Euler axis*). The 1st component of the quaternion (w) is typically the scalar value representing the rotation angle θ. The component formed of real and complex numbers ($x\ y\ z$) describes a unit vector representing the Euler axis. These four parameters, describing both angle and axis, enable quaternions to describe any rotation or orientation, which Euler angles fail to do. Another advantage of quaternions is that they do not require any form of trigonometric operation to describe orientation; these calculations demand a significant amount of computational resources. For example, it is possible to calculate the quaternion representing a rotation of a rigid body by simply multiplying the quaternion representation of both start and end positions. To determine the

Euler sequence describing this rotation, one cannot simply add or subtract corresponding axis rotations, as would be the case if the rotation only occurred on a single plane. The clear disadvantage to quaternions is that they are difficult to interpret and do not have an intuitive physical meaning. However, due to their efficiency and ability to avoid problems such as gimbal lock, quaternions have become the accepted method to represent 3D orientations in space for most computer graphic and robotic applications.

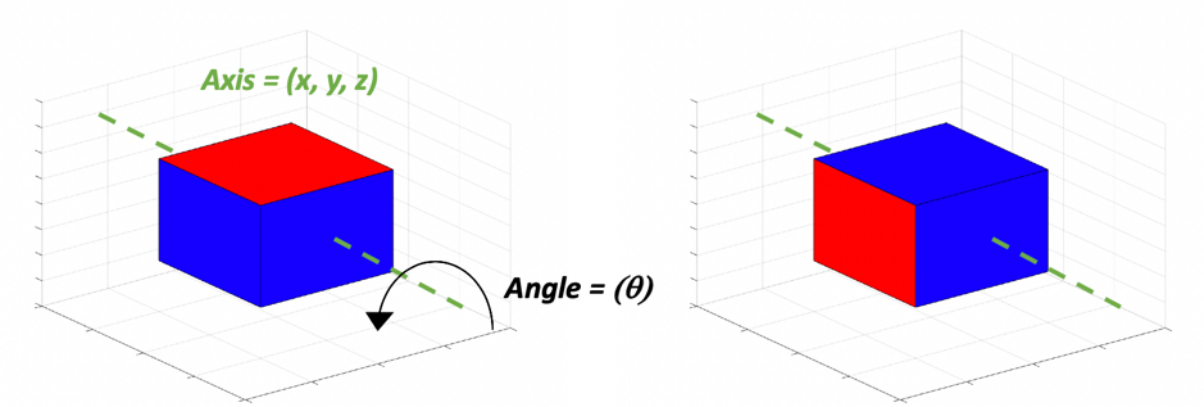


Figure 3.3. Any 3D rotation can be achieved with a single step rotation around the Euler axis.

3.1.3 Determining the accuracy of MIMU systems

For human joint kinematic assessment, it may be necessary to calculate relative body segment positions using two or more sensors. Therefore, when attempting to evaluate sensor accuracy, it is necessary to quantify how much a sensor's output differs to a validated reference system and to also quantify how much a sensor's output differs to the output of an additional sensor subject to the same movement conditions (102). These accuracy metrics will be referred to as *sensor-to-reference error* (SRE) and *sensor-to-sensor error* (SSE), respectively.

Several early studies provided valuable insights into the dynamic performance of inertial sensors under partially controlled conditions (103, 108-112). However, due to variations in experimental protocols, it was difficult to benchmark sensor and algorithm performance against one another. Additionally, with relatively short periods of data collection, it was not possible to evaluate the stability of orientation measurements over time. Recent controlled investigations under reproducible conditions have provided a more comprehensive understanding of MIMU error, with useful and generalisable findings. Lebel et al. assessed the impact of velocity and time on MIMU orientation error in three axes (101, 113). Using a motorized gimbal table, rotational movements of 90 °/s and 180 °/sec were generated, and position was referenced to an optical motion capture system. Root mean square (RMS) errors of 0.5° - 3.1° were reported for slow trials, but error deteriorated significantly for faster trials. Sensor accuracy was also shown to deteriorate significantly over time; error for some devices more than doubled within the one-minute data acquisition period. In all instances, SSE was worse than SRE. An embedded Kalman filter was used to derive orientation for all devices tested within the study.

Ricci et al. performed a detailed analysis of orientation error, assessing the impact of movement quality and filter choice (102). Using a robotic arm with inbuilt encoders for position reference, dynamic error was characterised under multiple movement conditions. Median values for SRE and SSE were reported in the range 0.3° - 4.5° and 0.4° - 3.5°, respectfully. Kalman and complementary class filters were both evaluated, demonstrating marginally better dynamic estimates with the Kalman filter. Dynamic settings were designed to represent normal human movement conditions; however, trials were only 60 seconds in length. Many clinical applications for inertial sensors, including those for assessment and monitoring, are likely to exceed this period. Currently, there is no published data that

characterises the accuracy of complimentary class filters over time. Additionally, there are no studies evaluating the Madgwick filter under clinically relevant movement conditions. As previously discussed, complimentary filters are easy to implement and require minimal computation time. These features make them a preferred choice for embedded systems and their performance should be better understood.

3.2 Study aims

The aim of this study was to characterise the orientation accuracy of a commercially available MIMU implementing the Madgwick orientation filter (103, 114). To realise this aim I needed to design and apply an off-line pre-processing pathway to raw MIMU sensor data to produce optimised time-synchronised nine-axis data for input to the orientation filter. Conditions during controlled dynamic testing were designed to represent a clinically relevant spectrum of human movement speeds and frequencies. The investigation provides a detailed assessment of the effects of time, movement speed and direction on sensor accuracy.

3.3 Methods

3.3.1 Data Pre-processing

Raw MIMU data requires preprocessing to create a consistently sampled and time-synchronized 9-axis dataset suitable for further processing and data analysis. These steps include resampling and interpolation of data, removal of high-frequency noise and correction of any sensor offset. All data processing steps were developed and performed in MATLAB (R2017a from MathWorks) and provided a foundation for all pre-processing required in further studies.

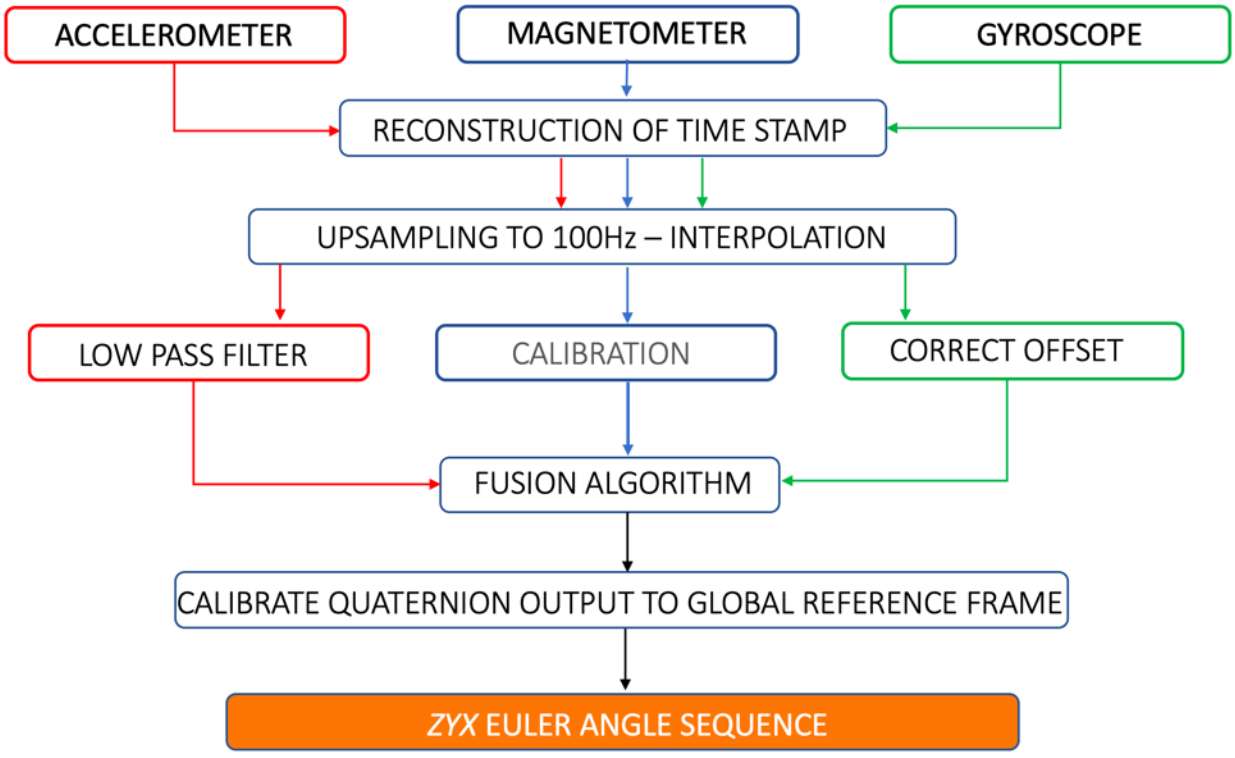


Figure 3.4. Overview of processing pipeline developed and tested within this study chapter and implemented in future work.

Re-sampling and Interpolation

The maximum permitted sampling frequency for each sensor modality was used for all studies: 100Hz Accelerometer, 100Hz Gyroscope, 25Hz Magnetometer. This was limited based on the transmission capacity of Bluetooth Low Energy (BLE) signaling, which was used for wireless data collection. To achieve a faithful recording of any signal, samples must be taken at a sufficient frequency. Nyquist's theorem suggests this sampling frequency ought to be at least twice the highest frequency component of what is being measured, to prevent a form of signal distortion known as aliasing (115). Considering typical human limb movement frequencies seen during ADLs or rehabilitation activity (116), the maximum sampling frequencies chosen were considered more than adequate.

Precise timestamps were available for all recorded data samples; however, the sampling rates provided by the accelerometer and gyroscope's internal clocks varied slightly and did not reliably produce 100 data points per second at corresponding one millisecond intervals. In order to produce data arrays at equal and consistently sampled rates, recorded data was up-sampled to an exact 100Hz for both the accelerometer and gyroscope. The magnetometer data was reliably recorded at 25Hz, with all data points 0.04 milliseconds apart. This data also required up-sampling to create a 100Hz data set corresponding with the other two sensor modalities. Up-sampling in both cases was achieved using a simple interpolation method based on the use of linear polynomials to create new data points within a range of a discrete set of known data points.

Filtering of Accelerometry Data

The random high-frequency noise contained within the accelerometer signal was removed by passing the data through a filter that retained only the low-frequency components of the signal. A low-pass Butterworth filter was used to achieve this, and a 10Hz cut off was applied to ensure no relevant movement frequencies were lost. Additionally, a bidirectional filtering technique was implemented to prevent a phase shift in the data (115).

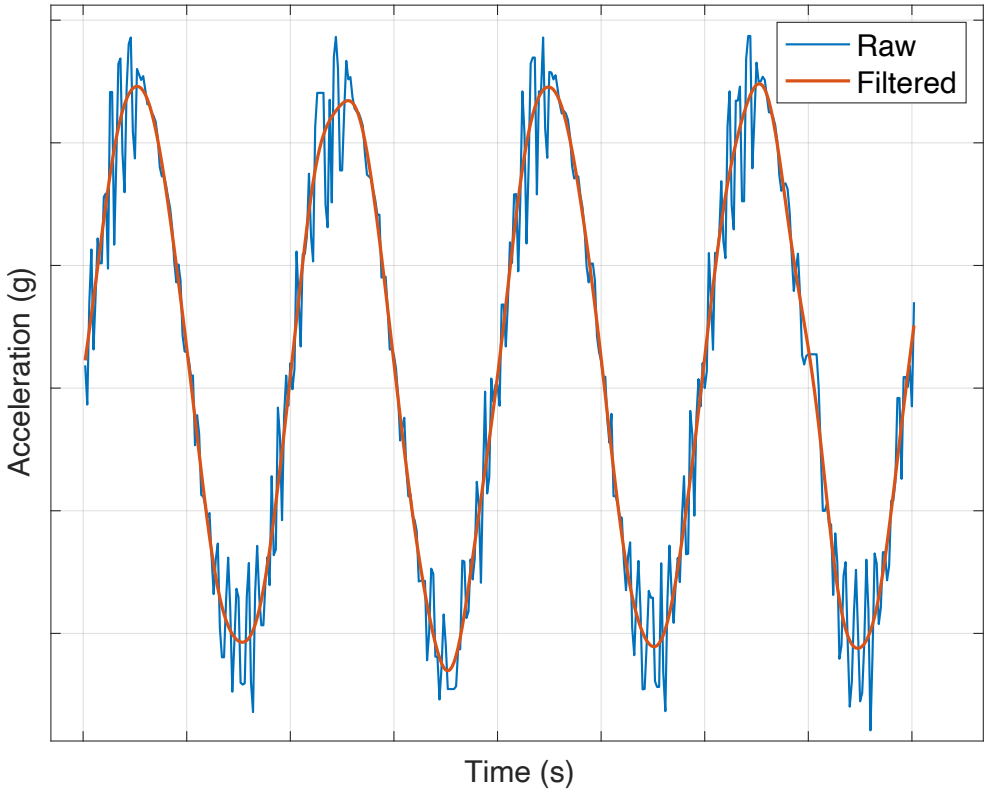


Figure 3.5. Raw and pre-processed accelerometer data after bidirectional low-pass filtering. Filtered data does not contain high-frequency noise and recorded events are not delayed relative to the original signal.

Sensor offset correction

Gyroscopes suffer from *bias*, which refers to an offset in the output signal from the true value, i.e. non-zero values are recorded when the gyroscope is stationary. The offset can be estimated by averaging the output value while the sensor is stationary and not undergoing any rotation or acceleration (117). Further dynamic measurements from all three axes within that specific data recording session can then be corrected using the measured offset.

The final output once all of these steps had been implemented were three-axis estimates for acceleration, angular rate and magnetic flux at consistent one millisecond intervals from the beginning of data collection (the coding pipeline is provided in Appendix D). This was subsequently used as input to the Madgwick orientation filter to generate orientation data in quaternions at a frequency of 100Hz. The version of the Madgwick filter implemented in this work was obtained from an open access code-sharing website (114).

3.3.2 Experimental setup

The setup consisted of two MIMUs (MetaMotion R, MbientLab, San Francisco, California, USA) mounted to a two-axis rotating platform (Tiab Ltd, Banbury UK) as shown in Figure 3.6. Each MIMU incorporates a tri-axis accelerometer (±2g, noise density:180μg/√Hz) gyroscope (±500°, noise density:0.008°/s/√Hz) and magnetometer (*x, y*-axis: ±1300μT, *z*-axis: ±2500μT, resolution: 0.3μT). Raw study data was streamed wirelessly via BLE to a smart device.

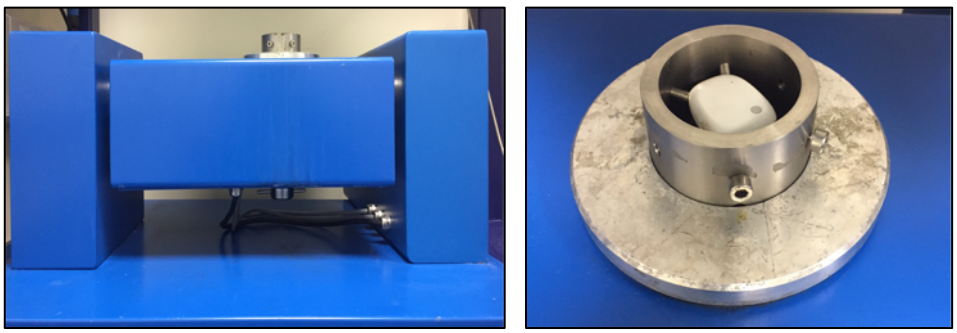


Figure 3.6. (a) Photograph of the oscillating platform from the front and a view of a sensor secured in the main housing. Two sensors could both be contained in the housing stacked on top of one another.

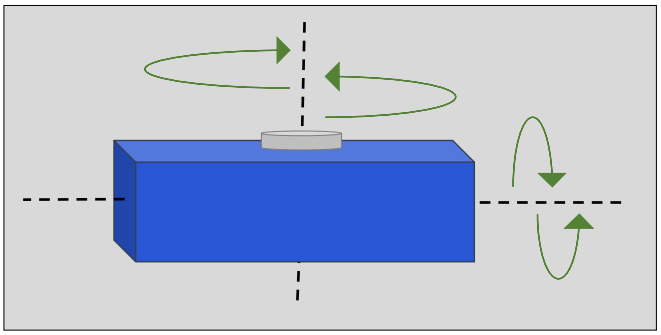


Figure 3.6. (b) Schematic diagram of the rotating platform, demonstrating the rotation possible in two axes. Encoders within the platform measure rotation relative to the starting orientation in both axes.

The rotating platform was controlled from a local computer generating sinusoidal movements, allowing the operator to control start position, frequency and amplitude of oscillations. The two MIMUs were rigidly housed within the platform, and it was assumed there was no movement between the two systems. Odometry of the platform was sampled at 100Hz, measured by digital encoders calibrated to ensure an orientation accuracy of 0.0068° (heading) and 0.0018° (attitude).

3.3.3 Study protocol

The protocol consisted of a series of static and dynamic movement trials. Dynamic trials were designed to test the accuracy of MIMU orientation output operating within a frequency bandwidth and movement speeds representative of typical of human limb motion performed within daily living activities (116, 118). Amplitude was tested at ten-degree intervals and limited to 50° due to safety constraints of the testing platform.

Three motionless trials lasting 60 minutes were performed to assess long-term static accuracy. Dynamic trials consisted of five minutes of continuous movement, initiated 20 seconds after activating the MIMUs. Every trial was initiated from a standardised position and performed three times. Orientation error for each five-minute trial was calculated through comparison with the reference platform encoder data. The MIMUs and the platform encoders operated on independent clocks, and precise alignment of the beginning of both dynamic outputs was achieved using a cross-correlation analysis. Attitude (horizontal) and heading (vertical) plane dynamic trials were conducted for all frequency/amplitude conditions. *Multi-axis* movement (simultaneous attitude and heading) trials were conducted for all amplitude settings at 1Hz. Considering two MIMU devices are tested in each condition, this provided 1,650 minutes of dynamic motion data for analysis.

The Madgwick algorithm employs a quaternion representation of orientation. To allow interpretation and standardised comparison with the reference encoder data, quaternion output from the filter was converted to Euler angles in a *ZYX* sequence and reported as attitude or heading. Horizontal movement accuracy was evaluated in a single plane only, as detection of *x* and *y* (*pitch* and *roll*) axis rotation utilises similar sensor modalities (accelerometer and

gyroscope) and undergo similar processing. The values for the adjustable parameter (*filter gain* "β") was derived from the maximum gyroscope error as recommended by Madgwick et al. (103).

The reference frame was defined by the average MIMU orientation calculated during a five-second static period preceding each trial. MIMU body frame rotation was calculated relative to the reference frame defined at the beginning of each trial. SRE was defined as the absolute difference in estimates of the orientation of the MIMU and the reference platform encoders at each sampled data point, set at one-millisecond intervals (30,000 data points per five-minute trial). Both SRE and SSE were calculated for all trials. SSE error was defined as the absolute difference between the outputs of two MIMUs undergoing the same movement.

The effect of time was evaluated using all attitude and heading trials performed at 1Hz. Error scores of *early* (T_1 = 0 - 150 seconds) and *late* (T_2 = 150 - 300 seconds) sections of the original 1Hz trials were compared to evaluate the impact of time on orientation error. Statistical description and tests implemented were selected considering the distribution of the data. All data processing and statistical analyses were performed in MATLAB (R2017a from MathWorks).

3.4 Results

Error values recorded throughout the dynamic trials were normally distributed (Figure 3.7) and absolute error values (considering the magnitude of error only) were positively skewed. A Kolmogorov-Smirnov test confirmed non-normality of the absolute error data;

therefore, the summary error for each trial is reported using median values. A summary data sheet containing median and 95% upper bounds is available in Appendix A.

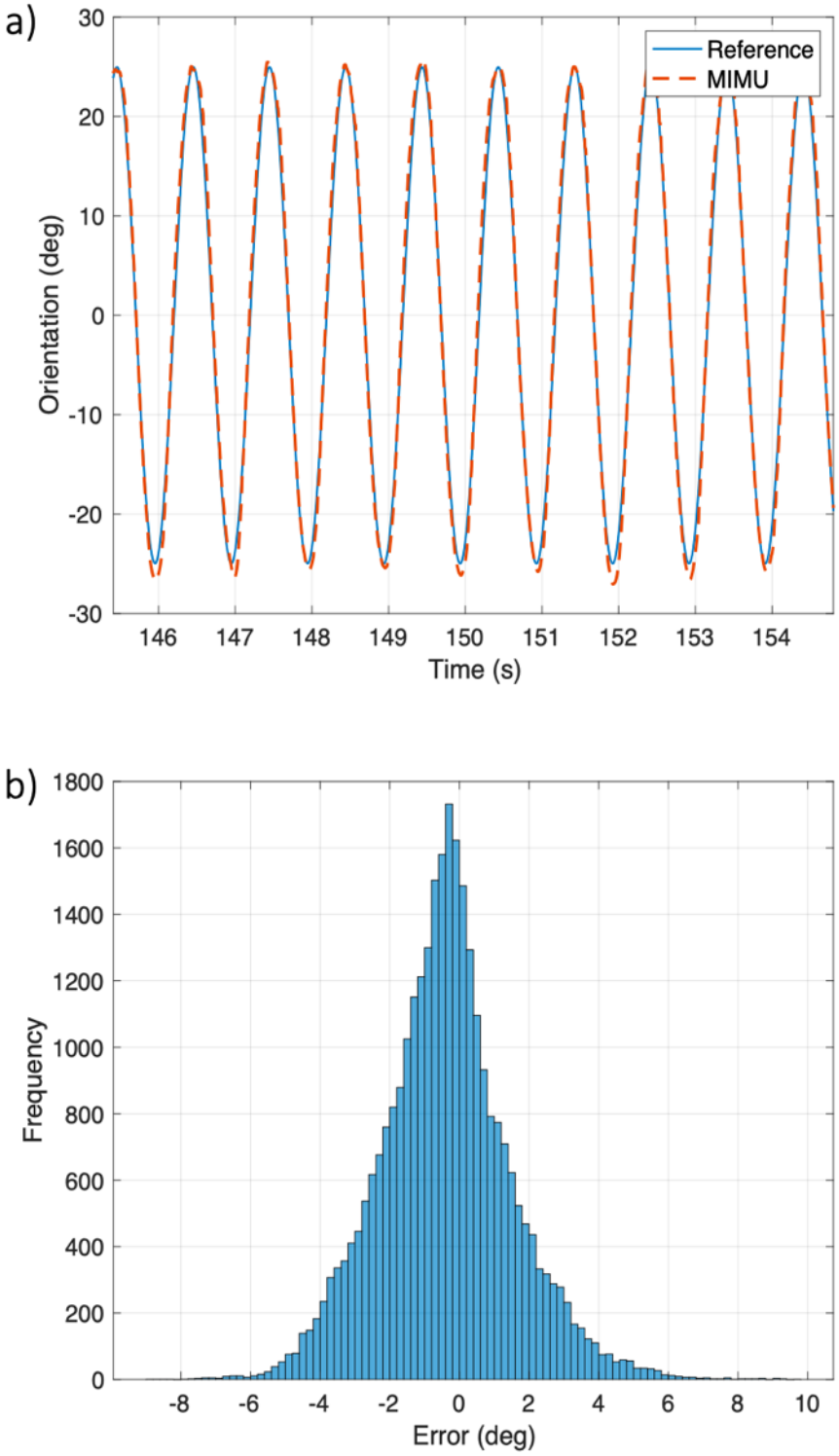


Figure 3.7. (a) MIMU orientation estimates compared with digital encoder data during section of dynamic attitude trial; (b) Normal distribution of recorded error in MIMU output the over entire five-minute trial.

The significance level for all tests was set to 0.05 and a Wilcoxon rank–sum test was used for the statistical comparison of two groups. Multiple groups were compared using a Kruskal–Wallis one–way ANOVA, followed by a single-step multiple comparison procedure using Tukey's method.

Median value SREs and SSEs varied from 0.17° to 7.54° and 0.16° to 14.62°, respectively. Statistical analysis showed SSE was significantly greater than absolute for both attitude and heading trials ($p < 0.001$). The direction of movement had an effect on sensor accuracy with attitude trials (0.17° to 4.57°) showing a better orientation accuracy than heading (0.43° to 7.54°), and this difference was also found to be statistically significant ($p < 0.001$).

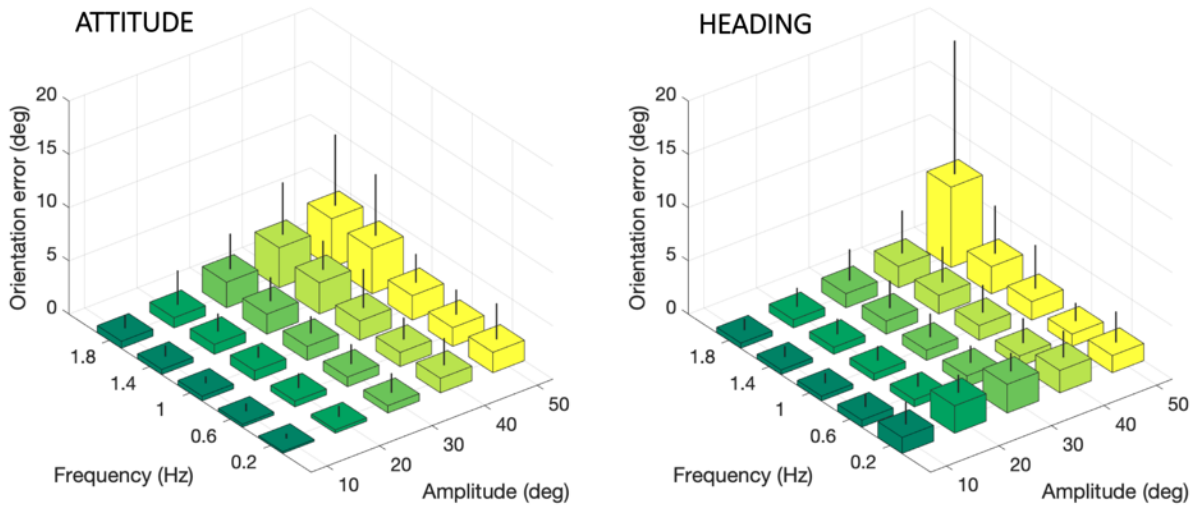


Figure 3.8. Bar charts displaying median SRE under different conditions in attitude and heading. Error bars display 95% upper limit.

Both movement amplitude and frequency had a statistically significant effect on orientation error ($p < 0.001$). Post-hoc comparisons demonstrate that all categories of trial within the amplitude and frequency range tested were significantly different from one another.

Multi-axis movement did not compromise accuracy in attitude estimation; however, accuracy in the heading plane was reduced ($p < 0.001$) when compared to estimates from the single-axis heading trials (Figure 3.9).

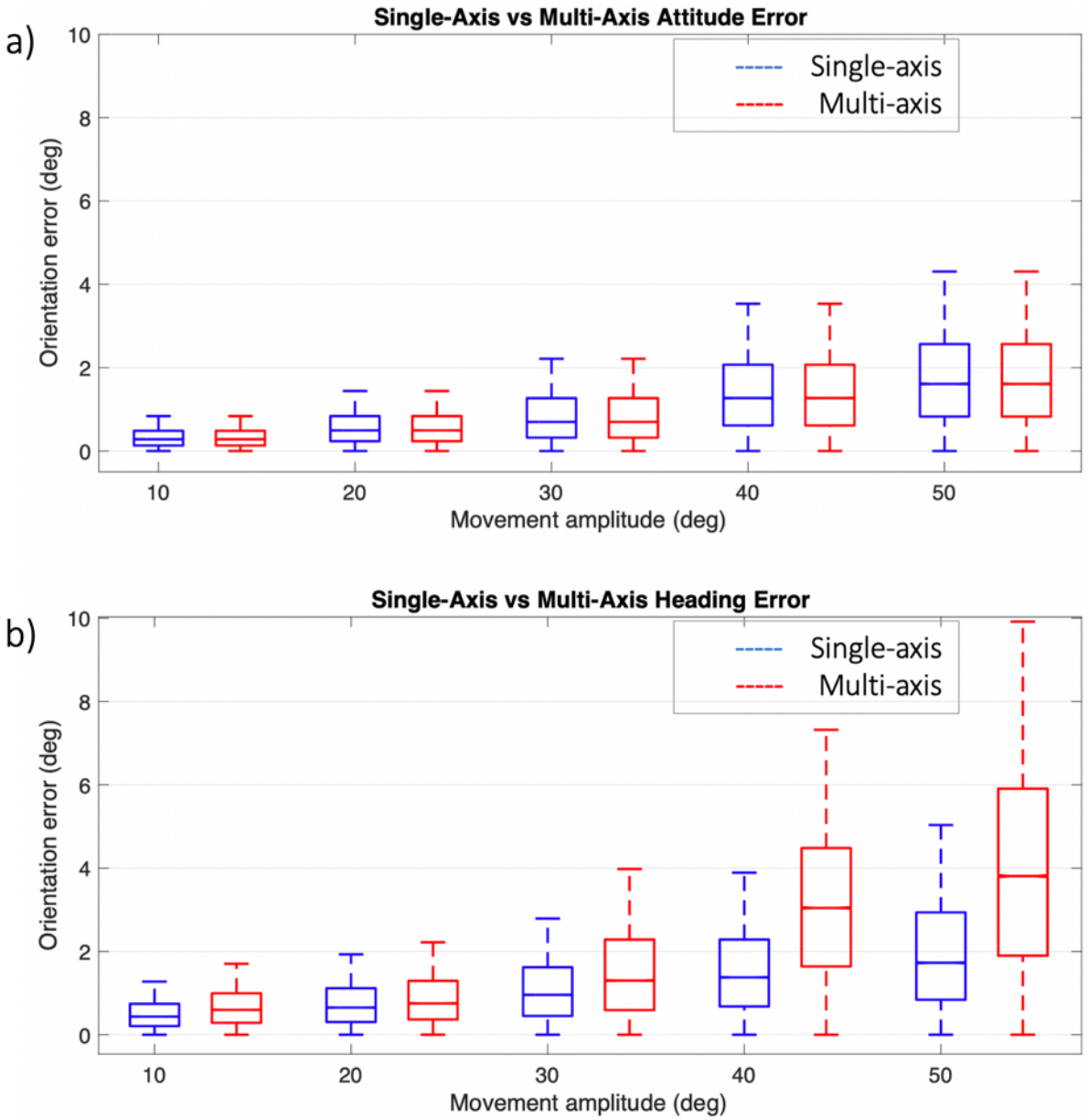


Figure 3.9. Box plots (centre bar: median; hinges: 1st & 3rd quartile; whiskers: 1.5 x interquartile range) comparing error during single-axis and multi-axis 1Hz trials for (a) attitude and (b) heading movement.

Accuracy deteriorated very slightly during the 1Hz trials, with individual trial differences in median error between T_1 and T_2 varying from - 0.08° to 0.15° (Figure 3.10). The overall change in accuracy over time for both attitude ($p < 0.001$) and heading ($p = 0.022$) trials were found to be statistically significant.

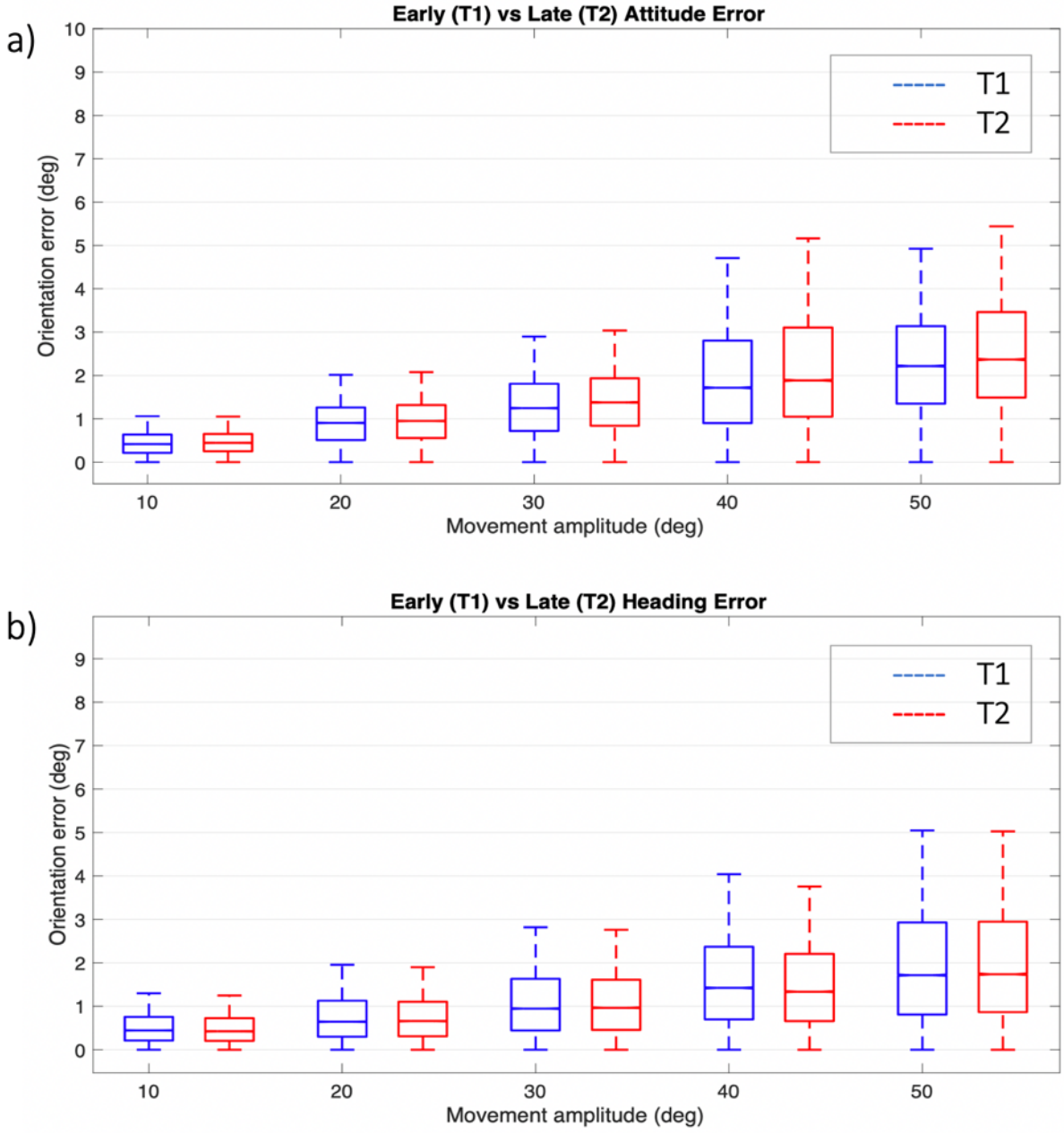


Figure 3.10. Box plots (centre bar: median; hinges: 1st & 3rd quartile; whiskers: 1.5 x interquartile range) comparing error during *Early* (T_1) and *Late* (T_2) portions of 1Hz trials for (a) attitude and (b) heading movement.

The MIMUs demonstrated stable orientation output under static conditions with median SRE in attitude and heading recorded as 0.03° (95% upper limit 0.16°) and 0.21° (95% upper limit 0.62°), respectively. Static SSE in attitude and heading planes were recorded as 0.05° and 0.24°, respectively.

3.5 Discussion

3.5.1 Implications of study

Measuring 3D orientation of upper-limb body segments with wearable sensors will allow for convenient assessment of joint angles and interpretation of shoulder movement during clinical activity. This study was designed to assess the static and dynamic accuracy of MIMU orientation estimates under controlled and repeatable conditions, eliminating error related to positioning, calibration and motion artefact. Dynamic assessment involved generating an automated oscillating movement and comparing MIMU estimates to data derived from calibrated encoders within the testing platform. The movement conditions were designed to simulate activity profiles expected in typical clinical applications and were also extended to extremes of movement rarely encountered in a clinical setting.

Oscillating movements at varying frequencies rather than continuous rotation was applied to better simulate human limb motion characteristics and test the filter's ability to respond to repetitive changes in speed and direction. The orientation filter employed in this study was chosen due to its reported computational efficiency and practicality of use, and this study provides the most comprehensive evaluation of this filter's performance. Existing studies

evaluating dynamic error of the Madgwick filter report RMS errors of up to 2.03°, although they do not characterize the spectrum of error or detail the factors influencing performance (103, 112, 119). This study demonstrates how sensor and filter performance is highly dependent on movement conditions, and the results are comparable with those of recent studies investigating alternative Kalman and complementary filters (102). The reduced processing time and lower sampling rate requirement demonstrate clear advantages of this method of filtering (119). While the demands for computational resources may not be noticeable in most computing environments, they can place a significant burden on more restricted embedded systems, such as those used for wearable platforms.

In keeping with previous findings, SSEs were higher than SREs, in both static and corresponding dynamic conditions (101, 102). While this metric does not inform us of true sensor orientation, it is essential for kinematic assessments requiring an estimate of relative body segment position. An observed source of systematic error within specific trials was imperfect time synchronisation between the two MIMUs (Figure 3.11). This specific source of error is not a factor when calculating SRE due to the cross-correlation method used to synchronise individual MIMU data to the encoder data. It is not clear whether this phenomenon has contributed to the relatively high SSE profiles reported in previous studies employing different sensor models and processing techniques.

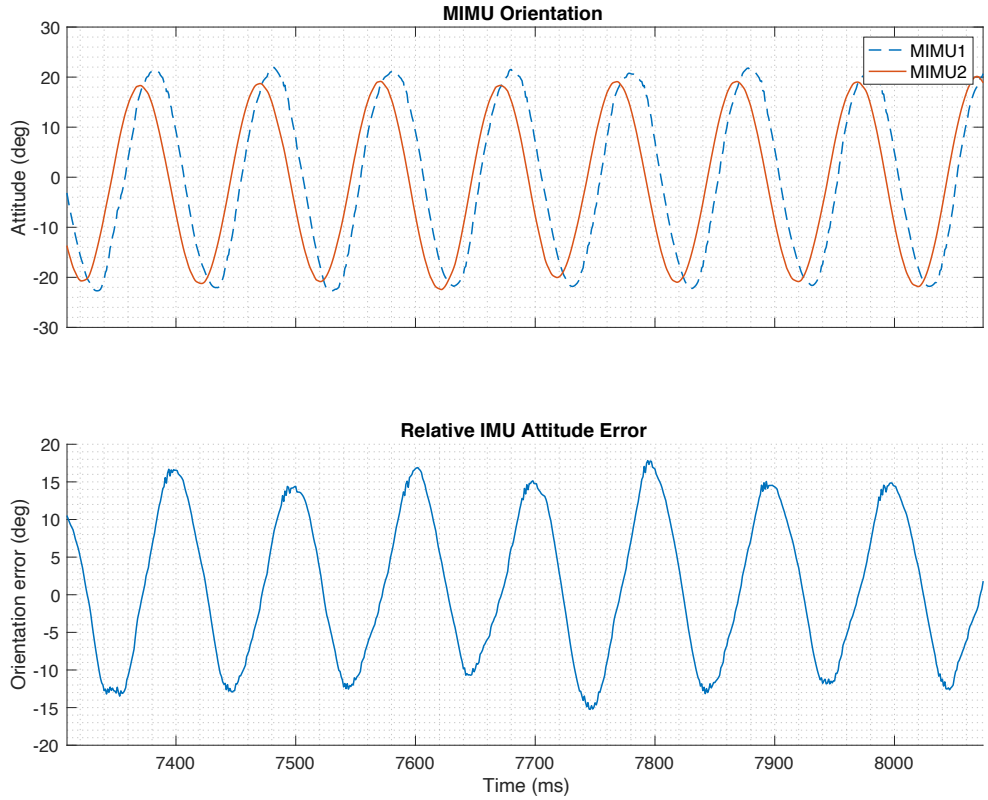


Figure 3.11. (a) Orientation trace of two MIMUs undergoing the same movement demonstrates time delay in MIMU1 output; (b) Resulting orientation error between both MIMU outputs.

It is possible that relative sensor error could be influenced by sensor positioning. Each trial was initiated with a standardised setup to minimise the effect of inertia on the trials repeatability; although, it was not possible to locate both sensors in the exact same position relative to the platform's centre of rotation. With each sensor being subject to slightly different inertial forces, systematic differences in individual sensor output may result. Finally, variations in the physical properties of individual sensor boards, imperfect mounting and thermal effects are factors known to influence the accuracy of inertial data acquisition (101, 112).

The higher error reported in heading plane trials, also consistent with previous reports, is due to the reliance on imperfect magnetometer readings to stabilise gyroscopic drifts in the

vertical axis (112). A variety of methods to mitigate magnetic distortion are proposed within the literature (120, 121) and have been shown to optimise heading estimates (119). The Madgwick filter incorporates magnetic distortion compensation to prevent attitude errors measured from the earth's magnetic field, however, heading estimates are unprotected. Feng et al. demonstrated the susceptibility of the Madgwick filter to magnetic disturbance, with heading estimates being compromised the most (119). Despite this relative weakness in filter performance, it must be noted that the majority of the heading trials within this study demonstrated median errors of less than 2°.

Through applying a variety of movement conditions, this study demonstrates a general increase in error associated with increasing frequency and amplitude. The only exception to this trend was the erroneously high error found in low speed heading trials. The 1Hz error profiles are comparable to previously reported performances of MIMUs implementing an embedded Kalman filter and an alternative complementary filter (Gauss-Newton algorithm), subject to similar motion conditions (Table 3.1) (102), suggesting all three methods would demonstrate a similar level of performance in routine clinical applications, e.g. patient kinematic assessments, rehabilitation activity recognition. The high attitude and heading error seen in 1.8Hz trials suggest the employed setup would be less suitable for more demanding physical applications, e.g. sporting assessments. Experimental conditions aimed at investigating such applications are required to explore this fully.

	Absolute Error (°)		Relative Error (°)	
Filter	**Attitude**	**Heading**	**Attitude**	**Heading**
Madgwick[1]	0.5 (1.4)	0.7 (1.9)	1.1 (6.3)	1.6 (5.3)
Kalman[2]	0.7 (1.9)	1.1 (4.1)	0.8 (1.9)	2.3 (8.2)
Complementary[2]	1.6 (4.0)	2.2 (4.9)	2.0 (3.9)	1.7 (4.5)

[1] 300sec of dynamic testing, sampling 100Hz. [2]Results from Ricci *et al.*: 20sec of dynamic testing sampling 128Hz (102).

Table 3.1. Comparison of median errors and 95% upper limits of Madgwick, Kalman and complementary filter. All dynamic trials involved 1.0Hz ±10° oscillatory movement.

Analysis of the 1Hz trials demonstrated some subtle deterioration in sensor accuracy over the five-minute trials, with the largest increase in absolute and relative median error recorded at 0.17° and 0.35°. Reliability in sensor readings over time is a necessity for MIMU technology aimed at long term patient monitoring. The observed change in accuracy was relatively small compared to investigations of sensors implementing Kalman filters, in some cases showing a marked deterioration in accuracy over a much shorter period of time [20]. This is due to the Kalman filter's accumulation of errors in the gyroscopic bias estimate (102, 113).

3.5.2 Study limitations

The results presented are dependent on the choice of sensor manufacturer, pre-processing methods and fusion algorithm parameters. Additionally, heading readings are likely to have been compromised by disturbances to the local magnetic field, and may not be representative of what could be achieved in magnetically *clean* environments. However, the findings do provide a reference range of what could be achieved in a *real-world* setting and allow us to evaluate the impact of movement quality on sensor accuracy. Due to equipment

constraints, the range of frequencies and amplitudes investigated in this study are not a complete representation of human movement profiles. It is important to apply caution when generalising these results to movement conditions that go beyond what is tested. However, the majority of rehabilitative and daily living tasks concentrate around 1Hz (116, 118).

Median of absolute error values was used to summarise and report MIMU orientation accuracy. A limitation of this error metric is the absence of information on systematic changes in accuracy related to velocity or amplitude within individual trials. The relative magnitude of peak amplitude errors for example, may not be accurately accounted for. However, experimental procedures and processing methods were standardised to allow reliable comparison and benchmarking of MIMU performance under different conditions.

3.5.3 Summary

This study reports acceptable accuracy and reliability of MIMU orientation output under controlled dynamic conditions representative of typical human movement. The performance of the Madgwick filter was comparable to reports of more established filtering methods, and these findings demonstrate a reliable influence of movement speed and direction on orientation error. The robustness of orientation accuracy over time and relative computational efficiency supports this form of filtering for real-time clinical applications. This processing pathway therefore forms the basis for MIMU data processing in the following chapters.

4 Assessment of Shoulder Range of Motion

Contents

4.1 Background 59
4.1.1 Considerations for 3D modelling of the shoulder 59
4.2 Study aims 62
4.3 Methods 63
4.3.1 Movement range calculation 63
4.3.2 Study participants and setup 64
4.3.3 Study protocol and statistical analysis 65
4.4 Results 66
4.5 Discussion 69
4.5.1 Implications of study 69
4.5.2 Study limitations 70
4.5.3 Summary 71

4.1 Background

Range of motion (RoM) of the upper limb is often restricted by subacromial shoulder pain, and this restriction can have a severe impact on one's ability to perform essential activities of daily living (ADLs) (122). Improving pain free RoM is an important goal for most shoulder rehabilitation strategies, and joint angles are often measured in clinical settings to evaluate function and progress with treatment (123). Clinical approaches to measuring joint movement, including the use of goniometry, are very examiner dependent and are limited by their subjectivity (124). Optical motion capture systems (OMCS), providing objective and reliable 3D capture of limb motion, are available in some specialist centres and can play an important role in the assessment and treatment of certain conditions (125, 126). Unfortunately, the use of these systems is restricted due to their limited availability, expense and impractical constraints (127). Wireless MIMU sensors can offer a practical and cost-effective solution to objective motion analysis. Most importantly, their portability also offers the potential to quantify joint movement in natural settings and conveniently track recovery throughout rehabilitation.

4.1.1 Considerations for 3D modelling of the shoulder

By attaching MIMU sensors to body segments, one is able to determine the orientation of those body segments relative to a reference frame based on the earth's gravitational and magnetic field. Relative orientation of the two adjacent body segments provides an estimation of a joint's RoM. Movement at the shoulder girdle occurs through multiple articulations, and it is virtually impossible to accurately model the movement of each individual segment with surface marker or wearable motion sensing technology. Therefore, most instrumented

measurements of shoulder RoM are in fact an evaluation of the humerothoracic joint: movement of the humerus relative to the thorax (59). This simplified approach to 3D modelling of the shoulder still presents significant usability issues, particularly from the perspective of an individual with shoulder pain. Referencing the position and movement of the torso requires a sensor to be secured firmly on the user's chest or back. Methods to achieve this, detailed within the literature and implemented in commercially available systems, include the use of bioadhesive patches, body straps or harnesses (100, 128, 129). These methods would undoubtedly impose a significant burden on most users, and are potentially unachievable if shoulder movement is restricted. The sensor-to-sensor error (SSE) analysis in the previous chapter also demonstrated how referencing sensor orientation from another active sensor can introduce error, particularly if sensor clocks are not perfectly synchronized.

Another practical limitation to 3D modelling of the shoulder joint with wearable devices, is the requirement for accurate *sensor-to-segment calibration* (63). The coordinate system used to define how the body segment moves, must be aligned with the coordinate system of the sensor to ensure meaningful and reliable output from the model. This calibration is achieved with an additional computational step; utilising positional information acquired from the sensor during the performance of selected static poses or functional movements. Naturally, if calibration procedures are not carried out correctly the model output is compromised (63), which places further responsibility on the user. It is unrealistic to expect patients to carry out complex calibration procedures correctly and consistently, and once again in some cases, this may not be achievable at all.

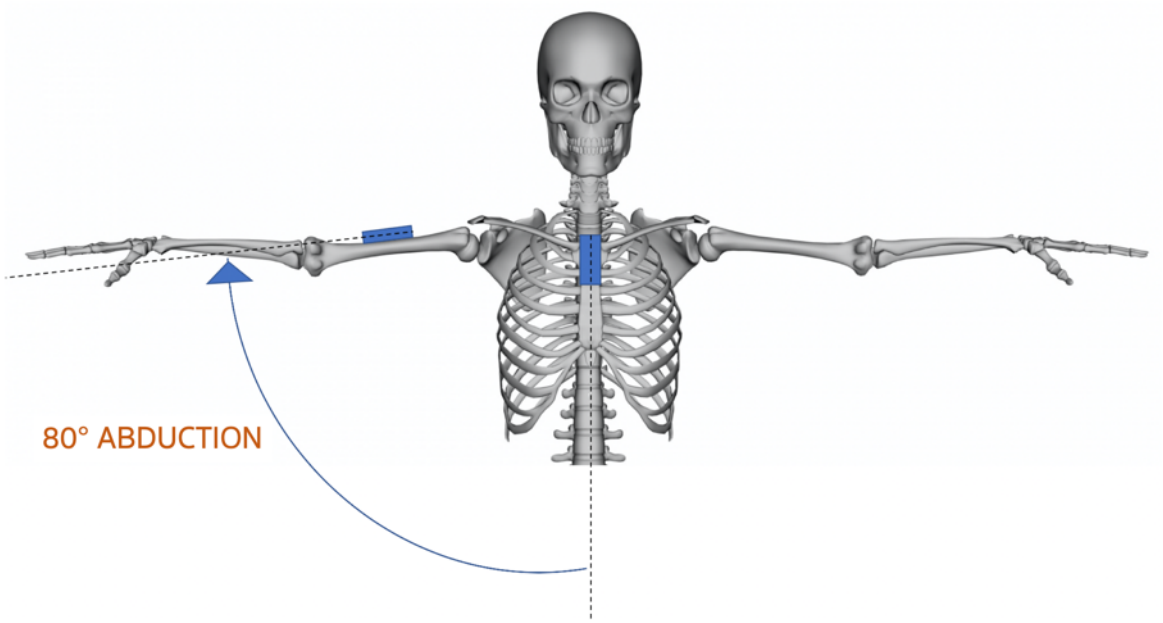


Figure 4.1. With 3D positional information of two sensors aligned to humerus and thorax, humerothoracic joint angles can be reported in three planes of movement, e.g. *abduction, flexion* and *rotation*.

To ensure the processes, as well as the sensor metrics, being investigated were clinically relevant and feasible in the context of patient rehabilitation, comprehensive modelling of the humero-thoracic joint requiring multiple sensors and detailed calibration was not considered. Rather, RoM was calculated as a relative value intended to quantify the magnitude of motion of the upper-arm segment during movement. This was achievable with a single sensor and not dependent on positioning of the device relative to the body segment; provided that position was consistent throughout movement. Motion artefact and body movement relative to the reference frame will inevitably affect measurement accuracy. Provided these factors had a consistent impact on MIMU measurement; the relative RoM estimate would provide a reliable, objective and practical tool for quantifying shoulder movement.

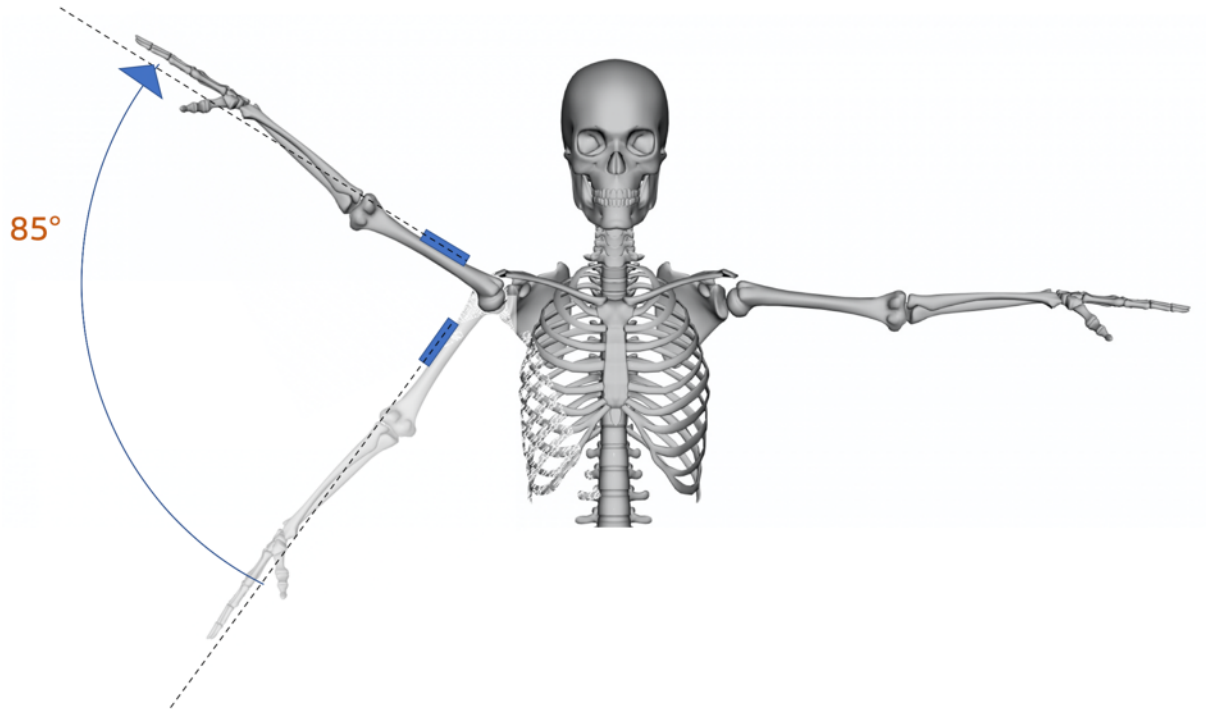


Figure 4.2. Relative movement of the upper arm segment can be calculated from sensor orientation data at two different positions. Alignment of coordinate systems are not required.

4.2 Study aims

This study investigated relative shoulder RoM estimates produced from a single wireless MIMU device mounted onto an ergonomically designed wearable accessory positioned on the upper arm. Both accuracy and precision of MIMU output during shoulder movement in three planes of motion were evaluated through comparison with a gold standard OMCS.

4.3 Methods

4.3.1 Movement range calculation

Raw nine-axis motion data (tri-axis accelerometer: ±2g, noise density:180μg/√Hz; tri-axis gyroscope: ±500°, noise density:0.008°/s/√Hz) sampled at 100Hz, 100Hz and 25Hz was acquired from an arm sensor. Raw sensor data underwent the pre-processing pipeline presented and tested in Chapter 3, and orientation of the sensor was delivered at 100Hz using the Madgwick filter.

The extreme positions (e.g. maximum abduction) on both MIMU and reference data streams were identified using a *peak analysis* function in MATLAB. RoM was defined as the relative difference in 3D orientation of the upper arm segment at both extremes of each movement. This was calculated by first determining the product of the rotation quaternions (q^A and q^B) representing each extreme position. The four-unit vector [$w\ x\ y\ z$] that results from the matrix multiplication is a quaternion (q^{AB}) representing the angle (θ in radians) and the axis for a rigid body to rotate from the positions represented by q^A to q^B. The scalar component of the quaternion (w) is equal to cos(θ/2), therefore θ = 2*acos(w). The θ value was subsequently converted from radians to degrees for analysis and reporting. Alignment with a specific reference frame was not necessary as only one sensor was being used and sensor rotation was always being referenced from a previous position registered within the same coordinate system.

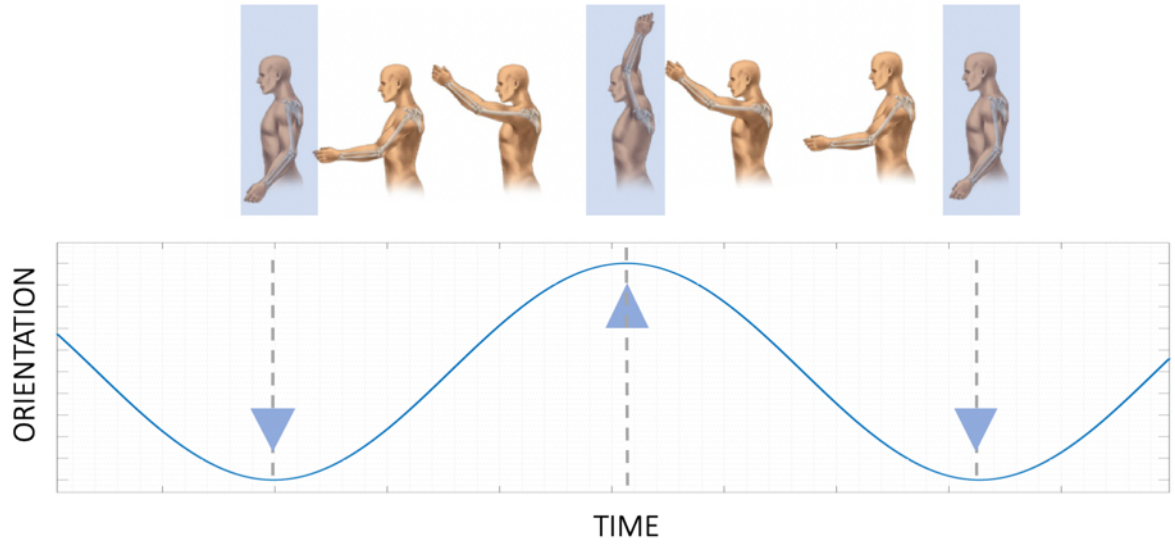


Figure 4.3. Peak analysis of data streams was performed using the dominant axis of the Euler angle trace.

4.3.2 Study participants and setup

Ethical approval for the study was granted by a UK National Health Service Research Ethics Committee (Appendix B). Twenty volunteers (mean age 29.6 [SD 4.5] years, 6 female, 18 right arm dominant, mean BMI 23.5 [SD 4.3]) participated in the study. All participants provided informed consent for their involvement in the study and denied experiencing any symptoms of shoulder pain or stiffness. Data collection took place in the Oxford Gait Laboratory to enable simultaneous motion assessment with an OMCS (Vicon Motion Systems Ltd), which provided a reference to the MIMU system output. Participants were required to wear eight retroreflective body markers (9mm), positioned on the upper limb and torso, to allow 14 calibrated cameras to capture movements of the humerus and torso. A MIMU (MetaMotion R, MbientLab, San Francisco, California, USA) device was mounted to an arm sleeve worn above the level of the right elbow. The wearable accessory was positioned to

ensure the sensor was located at the posterior distal aspect of the humerus, as per International Society for Biomechanics (ISB) guidance (130).

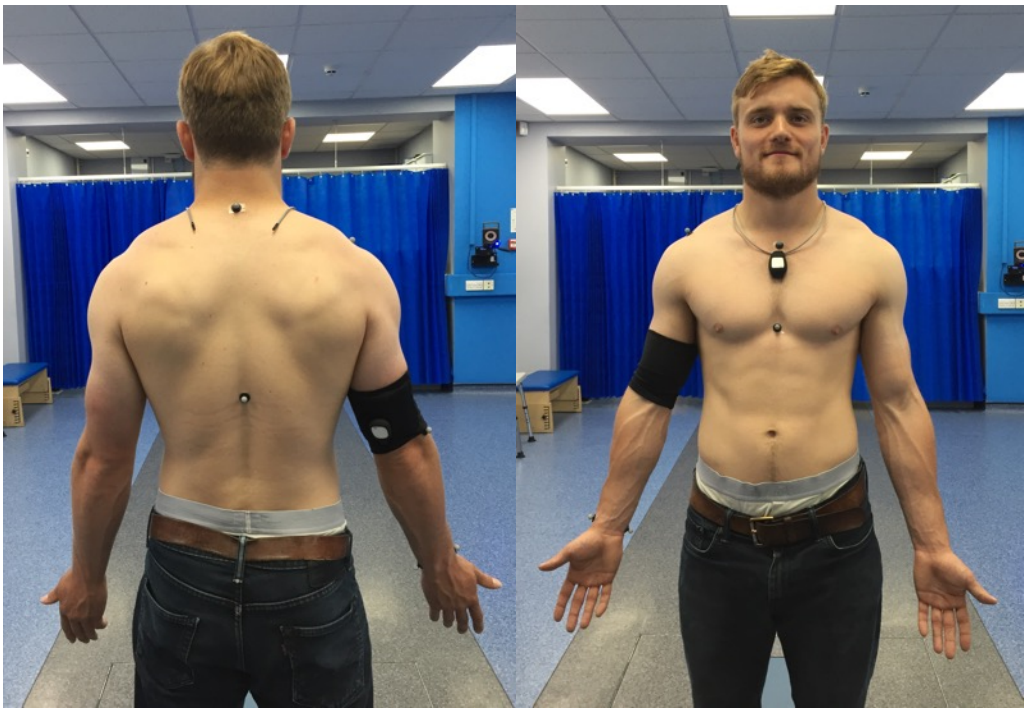


Figure 4.4. Study set up with wearable MIMU and retroreflective markers. A torso sensor was also trialled in several subjects however the data from this sensor was not considered for this analysis.

4.3.3 Study protocol and statistical analysis

The participants were instructed to perform 10 repetitions of three movements to evaluate shoulder motion in three DoF. The movements were shoulder abduction, forward flexion and rotation (with the arm abducted to 90°). MIMU sensor data was compared to the OMCS output in order to determine measurement accuracy. Upper arm segment orientation relative to a global reference frame was calculated using the Vicon plug in gait model. Setting the origin of the reference frame and calibration of the cameras was performed before data

collection using the standard recommended method by Vicon (131). Orientation output data was delivered at 100Hz and expressed as Euler angles. These values were subsequently converted to quaternions to allow calculation of the RoM metric described above.

The percentage difference in RoM estimates (MIMU and OMCS) for each repetition of movement was calculated, and values were averaged for each set of movements. Accuracy was also evaluated by quantifying humerothoracic waveform similarity of both systems using Pearson's correlation coefficient. The MIMU and the OMCS operated on independent clocks, and precise alignment of the beginning of both outputs was achieved using a cross-correlation analysis. The precision of MIMU RoM assessment was quantified by calculating intra-class correlation coefficient (ICC) estimates and their 95% confidence intervals for all measurements for each specific movement. ICCs were calculated based on a two-way, mixed-effects model testing for consistency across multiple measurements (132). The guideline for interpretation of ICC estimates by Koo and Li (2016) was used to determine if consistency was poor ($ICC < 0.50$), moderate ($0.50 \leq ICC < 0.75$), good ($0.75 \leq ICC < 0.9$) or excellent ($ICC \geq 0.9$) (133). All data processing and statistical analyses were performed in MATLAB (R2018a from MathWorks).

4.4 Results

All participants were able to carry out the required shoulder movements satisfactorily on the first attempt. No participants reported any discomfort or movement restriction related to the wearable accessory or skin markers. Data collected from two recordings were corrupted and could not be used for analysis. Table 4.1 displays the average differences between MIMU

and OMCS calculations of the RoM metric for three different movements. The measures of waveform similarity and score precision are also summarised in the table.

	RMS Difference	% Difference	Pearson's *r* Correlation Coefficient*	Intra-class Correlation Coefficient
Abduction	18.8° (15.0)	12.6 (10.2)	0.95	0.81 (0.74 - 0.86)
Flexion	23.7° (17.1)	15.1 (11.2)	0.87	0.91 (0.86 - 0.94)
Rotation	28.5° (15.1)	27.5 (12.8)	0.91	0.92 (0.89 - 0.94)

*Pearson *r* coefficients were statistically significant ($p < 0.05$) for all sessions

Table 4.1. Summary of results comparing MIMU and OMCS calculations of relative shoulder RoM.

RoM in all planes were underestimated by the MIMU system relative to the OMCS, with abduction plane movements being least affected and rotation plane movements being most affected. There was also variability across subjects in the extent to which RoM was underestimated by the MIMU system, with percentage differences varying from 5.5% to 42.0%. There was statistically significant consistency between both systems' measurements for each subject, quantified with the ICC estimates, which were considered good or excellent for all movements. Comparison of dominant axis waveform similarity between both systems demonstrated a statistically significant ($p < 0.5$) and strong correlation for all subjects performing all movements. Some data traces demonstrated a drift in reference frame (Figure 4.5); however, movement detection was consistent across both systems with respect to time.

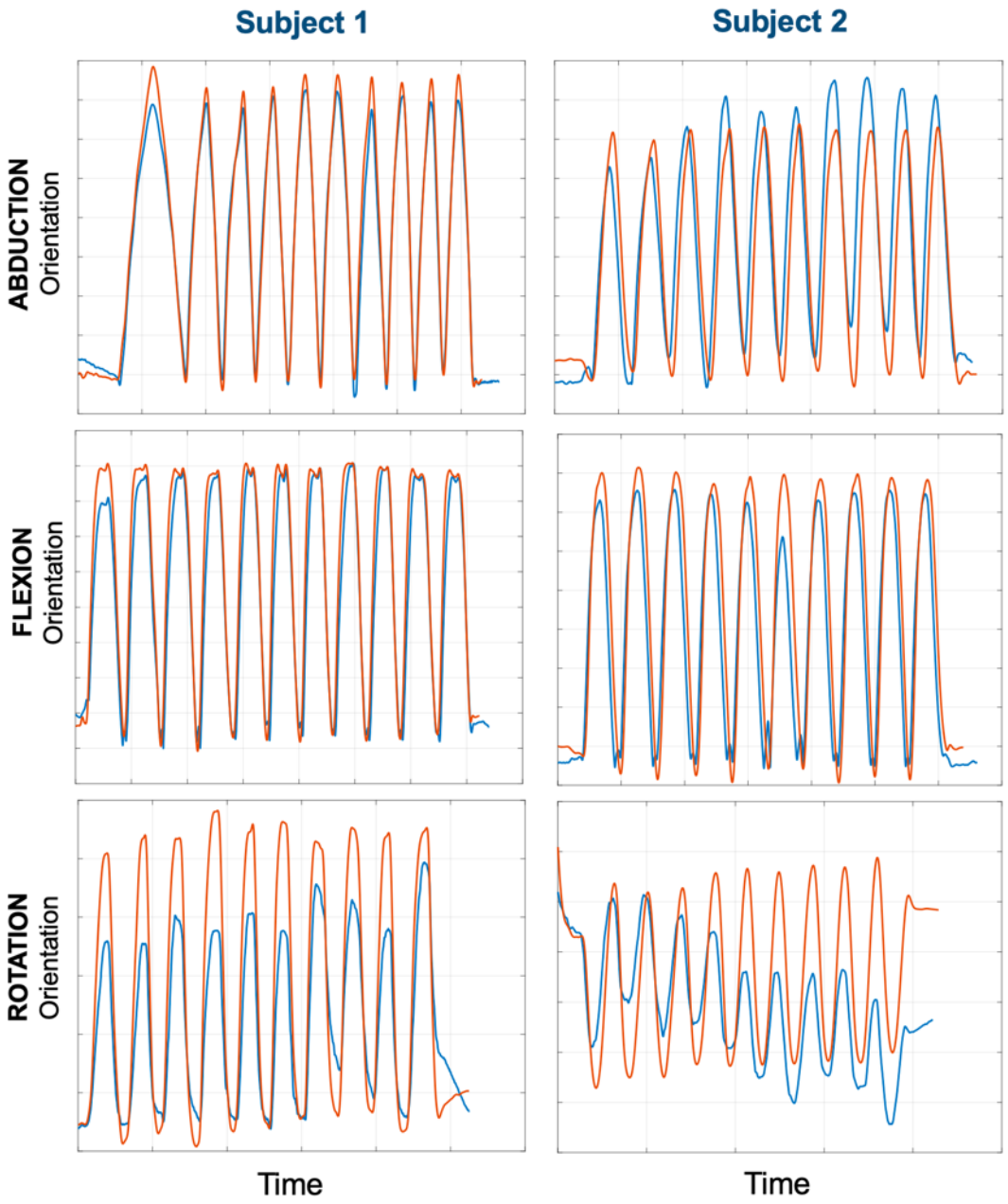


Figure 4.5. Single axis orientation of arm reported by MIMU sensor and OMCS over time for two subjects. MIMU rotation plane measurements are notably underestimated relative to abduction and flexion plane measurements. Drift in MIMU reference frame position demonstrated in abduction and rotation traces for Subject 2.

4.5 Discussion

4.5.1 Implications of study

The aim of this study was to evaluate the accuracy and precision of a single MIMU based wearable in quantifying RoM of the shoulder (humero-thoracic joint). An efficient and practical method was implemented to provide a single metric to quantify the magnitude of upper arm rotation during shoulder movements carried out on three orthogonal planes. When compared to a gold standard OMCS, the single sensor-based method underestimated the magnitude of motion. However, MIMU data traces did correlate very well with the reference data; demonstrating that sensor readings were reliable and responsive in tracking changes in direction and magnitude of movement. Analysis also confirmed that the magnitude of measurement error observed across measurements was consistent relative to the inherent variability for each subject.

The method tested in this study calculated the single rotation required to achieve a specific movement of the upper arm, using the upper arm's previous orientation as a reference. The metric is easy to interpret and can be applied to conveniently quantify the magnitude of shoulder movement performed during single plane shoulder rehabilitation exercises or other functional activities. It does not provide any interpretable information on the axis or trajectory of movement, and it is therefore not possible to ascertain from this metric alone if movements are carried out on the "correct*"* plane. This limitation was the compromise of adopting this low user-burden approach; involving only a single sensor and avoiding any form of complex sensor-to-segment calibration. The lack of a torso sensor also resulted in drifting of the sensor's

reference frame when subjects' bodies moved during motion capture. Calculating regular relative changes in movement rather than absolute position based on a predefined coordinate system, meant that this drift had a limited impact on measurement error or the consistency of readings throughout each trial.

4.5.2 Study limitations

A common problem shared by all forms of motion assessment based on surface mounted sensors or markers, is the artefact created by movement of the soft tissues relative to the underlying skeleton. Upper limb modelling using well established optical systems, including the Vicon Nexus Plug in gait model, are known to have inaccuracies related to soft tissue artefact and marker occlusion. OMCSs are regarded as the gold standard in clinical motion assessment; however, the imperfect nature of the study's reference system should be highlighted as a limitation (134).

Another important limitation of this study is the relatively small sample of subjects assessed. There was insufficient data to determine the possible impact of variables such as BMI on motion artefact and measurement error. It is also possible that the involvement of patients, with a different demographic profile to the healthy participants, could have resulted in different findings. Finally, identification of peak positions was a semi-automated process. While an automated peak analysis function was used, manual selection and extraction of relevant data streams was required. It was also necessary to adjust thresholds for analysis (e.g. distance between peaks, minimum peak height) between each subject. Successive studies will build on these methods to establish a fully-automated solution to segmenting motion data, and one that can be implemented in near real-time to support rehabilitation.

4.5.3 Summary

Measurement of shoulder movement using this MIMU based wearable system appeared to underestimate RoM relative to a gold standard OMCS, and this finding was consistent with previously published work. Considering the sensor accuracy reported in the last chapter, it was likely that the error in shoulder tracking was secondary to imperfect positioning of the sensor relative to the humerus throughout movement. Despite the effect of this motion artefact, wearable sensor measurements did demonstrate consistency, and could therefore provide a reliable method to characterise shoulder movement or quantify patient progress during a rehabilitation program. Extreme movement positions were determined with the assistance of the peak analysis function described in the methods. Moving forward, developments will have to made to the processing pipeline to allow data to be segmented automatically without any manual intervention. The development and testing of this code is described in Chapters 6 and 7, where the clinical value of MIMU derived orientation data is explored further. The following chapter continues to investigate MIMU derived shoulder kinematic assessment with MIMU sensors, exploring alternative motion parameters and their relevance to patient function and performance.

5 Instrumented Assessment of Shoulder Function

Contents

5.1 Introduction 73
5.1.1 Functional assessment in the clinical setting 73
5.1.2 Measuring clinical function with inertial sensors 74
5.1.3 Validation of clinical assessment tools 75
5.2 Study aims 76
5.3 Methods 77
5.3.1 Performance scores 77
5.3.2 Study participants 80
5.3.3 Assessment protocol 81
5.3.4 Statistical analysis 82
5.4 Results 84
5.4.1 Reliability 86
5.4.2 Known-group validity 86
5.4.3 Score responsiveness 88
5.4.4 Concurrent validity 89
5.5 Discussion 91
5.5.1 Implications of study 91
5.5.2 Study limitations 94
5.5.3 Summary 94

5.1 Introduction

5.1.1 Functional assessment in the clinical setting

Assessment of limb and joint function is an essential component of musculoskeletal practice and clinical decision-making relies on the accuracy and completeness of such assessments. Evaluation of treatment outcomes in the research setting also relies on standardised methods of assessment, and those aimed specifically at measuring functional domains typically lack objectivity or can be challenging to carry out. Clinical evaluation of shoulder range of motion and limb alignment provides little information on movement *quality* or capacity to carry out functional tasks. It is also known that the majority of clinical tests have multiple psychometric and performance deficiencies (135).

Validated patient reported outcome measures (PROMs) include measures of function, and are valuable in defining a patient's perspective of their clinical state. However, it is well established that these tools can suffer from ceiling effects and pain dominance can dramatically distort the functional assessment (136). For example, a reduction in pain after lower-limb joint replacement has been associated with patient reports of improvement in their functional ability, despite a significant increase in time to complete performance tasks (137). An additional drawback to the use of PROMs is that most systems were originally developed and validated for research purposes, and their extrapolation to clinical practice may not always be appropriate and valid, resulting in inaccurate interpretation of data (138).

5.1.2 Measuring clinical function with inertial sensors

Through human motion analysis, it is possible to quantify abnormal movement through measurement of kinematic data (Table 5.1). This information can be used to evaluate musculoskeletal conditions of the shoulder and can have implications on clinical practice (55). While there is great potential to gather objective and clinically relevant information, instrumented systems used for quantitative movement analysis, including electromagnetic, optical and mechanical techniques (50-54), are typically restricted to controlled laboratory settings, inaccessible to most clinicians due to cost, limited availability and the requirement for specific technical expertise. Self-contained wearable inertial sensors provide an opportunity to overcome these restrictions, offering practical alternatives to established methods of kinematic and functional assessment (59). Additionally, these sensor modalities offer the potential to calculate metrics representing a range of functional movement parameters, that go beyond established metrics such as range of movement (68). Quality of movement control, coordination and fluency of movement for example have been recognised as important components of physical function (56, 79), and practical tools to quantify these features may prove useful in the evaluation of treatment progress or outcomes.

Range of Movement	Peak range of movement
	Pain free range of movement
Speed	Task time
	Mean speed
	Maximum speed
Efficacy	Active movement index
Efficiency	Hand path ratio
	Index of performance
Accuracy	Movement deviation
	Target error
	Spatial overshoot
Smoothness	Mean/maximum velocity ratio
	Number of velocity peaks
	Mean arrest period ratio
	Jerk
	Accelerometer zero-crossing count
	Spectral arc length
Control Strategy	Time to peak velocity

Table 5.1. A range of end-point kinematic parameters classified by the movement characteristics they represent.

5.1.3 Validation of clinical assessment tools

Rigorous validation of a clinical assessment tool or measure is important to ensure the results of such assessments can be trusted. A valid tool is one that assesses what it claims to assess, and validity is a property of inferences made, supported by a spectrum of evidence. Validation does not give a simple yes or no outcome regarding trustworthiness of assessment

outcomes, but rather a judgement on trustworthiness depending on the intended application and context. A pre-requisite to a measure being valid is that it is reliable; first one needs to demonstrate that a test is reproducible by checking consistency across time and different observers (139).

In order to present a broad spectrum of validity evidence, several properties must be evaluated to demonstrate the relationship between a test's content and the construct it is intended to measure. This can be achieved through convergent and known-group validity (140). Convergent validity is commonly regarded as an important component of construct validity and involves demonstrating that a measure is related to one that it should relate to (i.e. a measure that is known to represent the construct). Scores from instruments that indicate significant differences in groups that would be expected to differ in their presentation, provides known-group validity. Responsiveness refers to the extent to which a measurement tool is able to detect relevant change when it has occurred. In the context of a clinical assessment tool, this is critically important particularly if treatment outcomes are to be evaluated (139).

5.2 Study aims

This study aimed to evaluate the psychometric properties of certain kinematic scores intended to reflect clinically relevant domains of upper limb function. To achieve this, I had to establish methods to calculate and summarise a selection of metrics from raw inertial sensor data. Data was collected from subjects with pain-free shoulders and a cohort of patients experiencing subacromial pain, with a view to exploring score reliability, responsiveness and known-group validity and convergent validity.

5.3 Methods

5.3.1 Performance scores

Only data from the accelerometer and gyroscope were required for this study. The raw six-axis sensor data (tri-axis accelerometer: ±2g, noise density:180μg/√Hz; tri-axis gyroscope: ±500°, noise density:0.008°/s/√Hz) sampled at 100Hz was acquired from two inertial sensors (MetaMotion R, MbientLab, San Francisco, California, USA) attached to both arms with a bioadhesive patch. Data was collected wirelessly via Bluetooth Low Energy to a smart device and processing was performed offline. Pre-processing of raw accelerometer and gyroscope data was performed using steps described in Chapter 3. The subsequent calculations to derive end-point performance variables and justification for their use are discussed below.

Movement *smoothness* was characterised using the SPARC (modified spectral arc length) metric (141). This method involves measuring the arc length of the Fourier magnitude spectrum to quantify intermittency in movement. A key advantage of this metric over alternatives listed in Table 5.1 is the lack of dependence on task duration and movement amplitude, potentially improving the generalisability of the score. SPARC values were calculated from pre-processed gyroscope data using a MATLAB function produced by the authors of the manuscript first describing the method. This code was available from an open access repository within GitHub, an online code hosting and sharing platform (142). The score used for this study was based on the average of SPARC values calculated from angular velocity (degrees per second) data of all three gyroscope axes (*x, y* and *z*) for each trial.

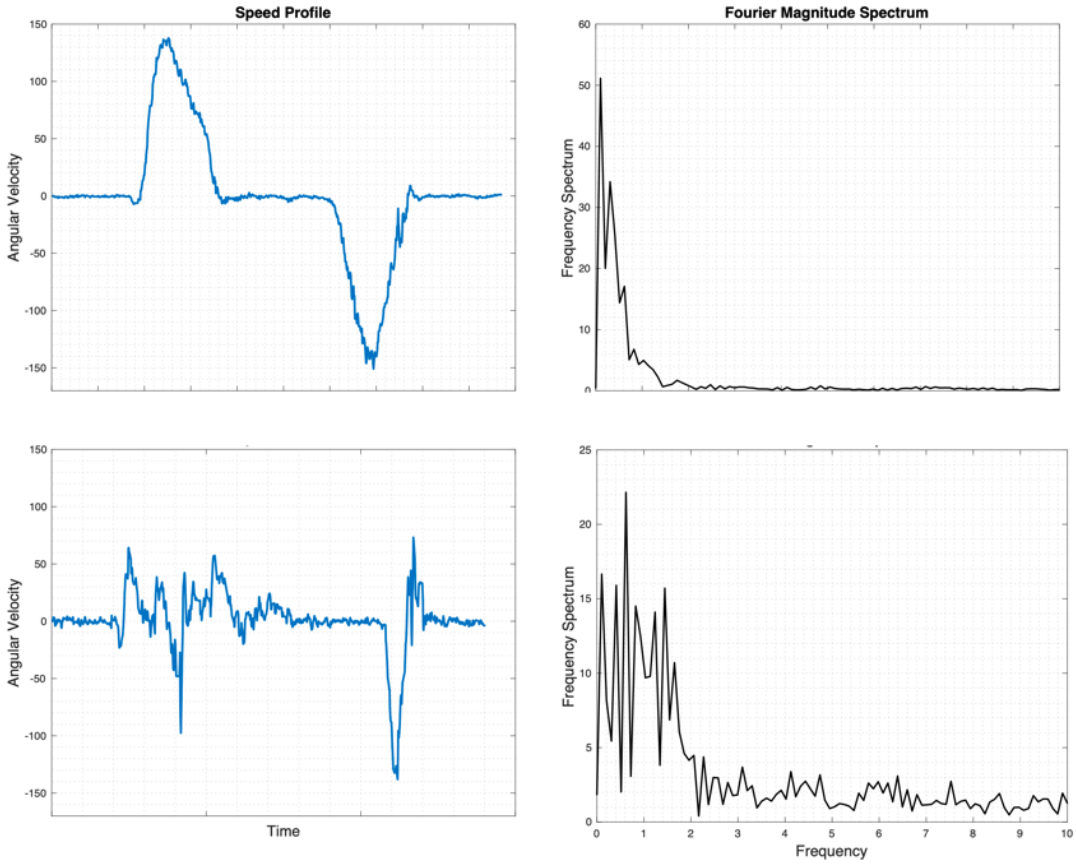


Figure 5.1. Single-axis angular velocity profiles (left column) during shoulder elevation and the corresponding Fourier magnitude spectra (right column). The bottom row profile displays reduced smoothness with greater number of submovements. The increase in submovements is reflected in the complexity of the frequency spectrum.

A measure of movement *speed* was calculated based on peak angular velocity (degrees per second) recorded by the gyroscope. The speed metric in this study is not an average value calculated from the gyroscope output for the duration of the trial, as this would naturally have a very high correlation with task duration. Peak speed is the metric being investigated and several methods have been used to represent this value, including measuring peak-to-peak differences in the angular rate signal, averaged from the three gyroscope axes. In this study the speed metric is represented by the 99th centile of normalised magnitude vector values of the tri-axis gyroscope. This is based on the

better intrasession agreement and test-retest reliability demonstrated in preliminary work.

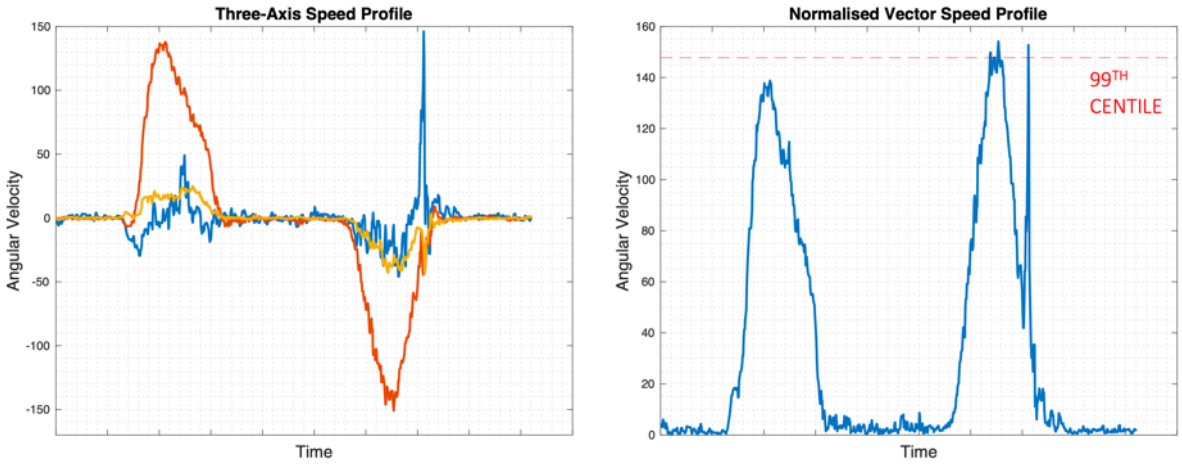


Figure 5.2. Tri-axis angular velocity profile (left) during shoulder elevation and plot of corresponding normalised vector (right).

The final inertial score calculated was a metric referred to as *power*. This score is simply the product of the peak angular velocity (degrees per second) and peak acceleration (g) during a trial. This is a similar concept to the *P* score proposed by Coley et al (68), except both representations of velocity and acceleration are derived from the 99th centile of normalised magnitude vector values. Once again, these values demonstrated better consistency in preliminary work when compared to scores calculated from individual axis data.

Time to complete task (seconds) was calculated in addition to benchmark inertial scores against a practical and established performance indicator.

Raw data pre-processing involved low-pass filtering of the accelerometer data with a 6th Order Butterworth filter (10Hz cut-off). Data streams were segmented manually following

visual inspection of the accelerometer data trace. Segmentation points corresponded to static positioning of the subject's arm between each movement, which could be identified on the accelerometer trace. Scores were calculated for each segment and individual summary scores for both the *back* and *bulb* component were reported as averages of the three repetitions of each movement for each shoulder within a single trial. Scores representing each shoulder, as well as asymmetry scores, were calculated. Asymmetry scores are relative scores that express the performance of the affected shoulder as a percentage of the performance of the healthy shoulder and were calculated as following:

$$Asymmetry\ Score = \frac{Score_{healthy\ shoulder} - Score_{painful\ shoulder}}{Score_{healthy\ shoulder}} \times 100$$

For healthy control subjects, asymmetry scores were calculated with the *healthy* and *painful* shoulder scores representing the dominant and non-dominant shoulder score, respectively. Endpoint performance variables calculated on all shoulders included the following:

5.3.2 Study participants

Ethical approval for the study was granted by a UK National Health Service Research Ethics Committee (Appendix B). Twenty-five patients (mean age 58.9 [SD 10.5] years, 13 female, 22 right arm dominant, mean BMI 28.5 [SD 6.3]) with a diagnosis of unilateral subacromial shoulder pain participated in this study. Diagnoses were made by an orthopaedic shoulder specialist and eligible patients were identified on referral to the physiotherapy department of an orthopaedic hospital (Nuffield Orthopaedic Centre, Oxford, UK). All patients

were scheduled to undergo a course of shoulder rehabilitation involving sessions with a physiotherapist and prescribed home exercises. Criteria for exclusion within the patient group included bilateral shoulder pathology, adhesive capsulitis, inflammatory arthropathy, any neuromuscular condition, cervical spondylosis or a history of recent trauma. Seven patients had partial or full degenerative rotator cuff tears and 21 patients had previously received treatment in the form of physiotherapy or steroid injections. 50 control subjects (mean age 48.4 [SD 21.2] years, 21 female, 42 right arm dominant, mean BMI 24.3 [SD 2.7]) reporting not to experience any shoulder pain or dysfunction were recruited to provide a comparative data set. All subjects meeting the eligibility criteria provided informed consent for their participation in this study.

5.3.3 Assessment protocol

Assessments involved all participants carrying out the *back to bulb* procedure, first described by Pichonnaz et al. (143). The procedure is comprised of two functional movements (internally rotating shoulder to bring the hand to the lower back and elevating shoulder to bring the hand to position in front and above the head) repeated three times at a self-selected pace (Figure 5.1). Both movements are considered to reflect upper limb functioning across a range of important daily living activities.

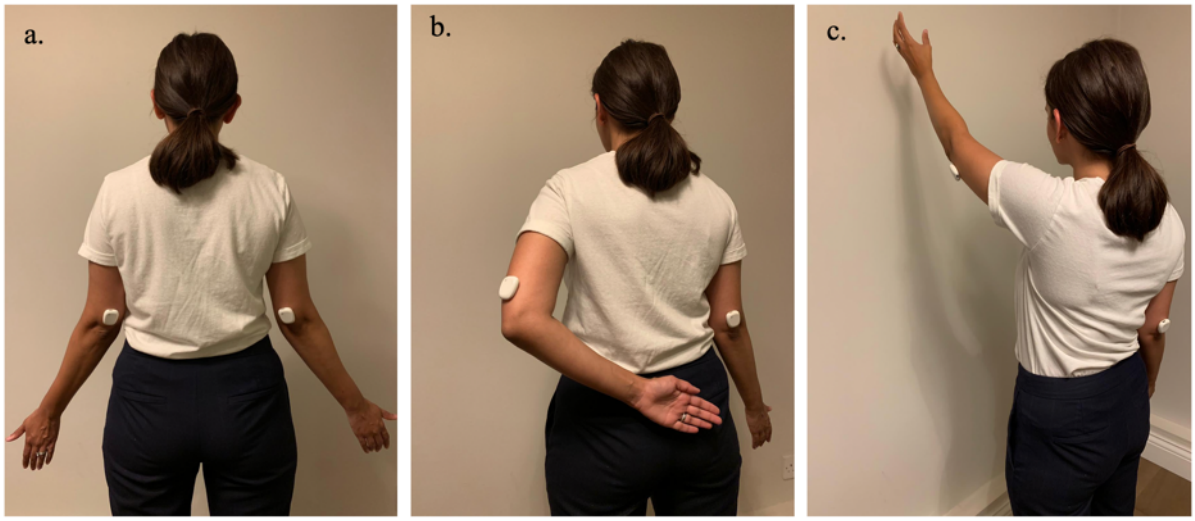


Figure 5.3. Inertial sensor setup with (a) start and target end positions for (b) back and (c) bulb movements. Procedure performed at a self-selected pace and repeated three times for both arms.

Two assessments were performed on all patients: at baseline and follow up review 92.0 days (SD 17.3) after initiation of treatment. Single assessments were performed on all healthy subjects, except for ten individuals who underwent assessments by two investigators to evaluate inter-rater reliability. In addition to the *back to bulb* procedure, all participants were asked to complete an Oxford Shoulder Score (OSS), which is a validated PROM questionnaire for the assessment of shoulder pain and function (144).

5.3.4 Statistical analysis

Descriptive statistics were calculated for demographic information, clinical and inertial scores at baseline and follow-up. The Kolmogorov-Smirnoff test was performed to confirm the normality of the data and parametric statistical analyses were performed. Tests to evaluate group differences were carried out at the 0.05 significance level. Multiple linear regression

models were applied to check for score dependence on the following variables – age, BMI, gender and handedness.

Reliability testing was performed using individual shoulder scores within both patient and volunteer groups. Inter-rater reliability was evaluated using data from 20 shoulders of ten healthy subjects; two sets of individual shoulder scores were generated during different sessions with a different investigator. Intra-rater reliability was evaluated using all patient healthy shoulder data collected at baseline and follow-up assessment. Intra-class correlation coefficient (ICC) estimates and their 95% confidence intervals were calculated based on a two-way, random-effects model testing for absolute agreement (132). The guideline for interpretation of ICC estimates by Koo and Li (2016) was used to determine if reliability was poor ($ICC < 0.50$), moderate ($0.50 \leq ICC < 0.75$), good ($0.75 \leq ICC < 0.9$) or excellent ($ICC \geq 0.9$) (133).

Known-group validity was assessed by drawing comparisons between study groups using parametric tests to evaluate similarity in score profiles. Receiver operator characteristic (ROC) curve analyses were carried out and the area under curve (AUC) was used to quantify diagnostic power. The responsiveness of scores to detect changes over time was evaluated by paired comparison of baseline scores with those following treatment, with calculation of Cohen's *d* effect sizes with 95% confidence intervals. The guideline for interpretation of effect sizes by Cohen (1988) was used to determine if effect sizes were poor ($d \geq 0.2$), medium ($d \geq 0.5$) or large ($d \geq 0.8$). Pearson correlation coefficients were used to determine strength of relationship between individual scores and the OSS. Data processing, analytics and statistical analysis were performed in MATLAB (R2018b from MathWorks).

5.4 Results

All participants were able to complete the test satisfactorily on first attempt. No complications were encountered, although some patients did report mild to moderate discomfort at the extremes of movement for both parts of the procedure. Multiple linear regression analysis identified age to be a significant predictor for all inertial scores. Increased age was associated with reduced smoothness (β = -0.03, p = 0.009, R^2 = 0.16) and increased speed (β = 0.99, p = 0.003, R^2 = 0.25) and power (β = 1.81, p < 0.001, R^2 = 0.24) scores. No significant impact was found from hand dominance, gender or BMI. Random sampling of healthy control subjects within age brackets corresponding to the patient group was performed. Data from an aged matched sub-cohort of 25 control subjects (mean age 59.1 SD 15.5 years, 12 female, 20 right arm dominant, mean BMI 24.9 SD 2.7) was used for all group analyses to control for the confirmed effect of age on score value. Matched control group and patient group baseline and follow-up summary scores are presented in Table 5.2.

			Matched Control		Patient Baseline		Patient Follow-Up		Effect Size
		Age	59.1 (15.5)		58.9 (10.5)				
		Gender	12 F 13 M		13 F 12 M				
		BMI	24.9 (2.7)		28.5 (6.3)				
		Dominance	20 R 5 L		22 R 3 L				
		OSS	48 (0)		33.9 (7.2)		38.6 (5.8)		0.70 (0.28 – 1.20)
Back	Individual	Time	5.2 (1.1)	5.0 (1.3)	8.5 (2.0)	7.0 (1.0)	6.5 (1.1)	6.5 (1.0)	1.21 (0.83 - 1.68)
		Smoothness	- 9.8 (2.6)	- 9.8 (3.1)	- 12.4 (3.5)	- 11.6 (2.1)	- 10.6 (1.7)	- 11.2 (2.2)	0.64 (0.21 - 1.06)
		Speed	274.6 (97.4)	281.3 (107.4)	153.5 (61.2)	207.7 (45.5)	204.5 (51.3)	217.0 (40.9)	0.88 (0.59 - 1.32)
		Power	337.9 (141.5)	344.9 (165.6)	179.0 (78.6)	249.1 (63.6)	248.2 (78.2)	260.1 (106.8)	0.86 (0.60 - 1.27)
	Asymmetry	Time	3.5 (13.8)		29.8 (36.0)		- 0.7 (19.3)		1.03 (0.59 - 1.60)
		Smoothness	- 3.1 (25.5)		- 8.2 (27.2)		3.4 (18.5)		0.49 (0.01 - 0.99)
		Speed	- 0.5 (19.8)		25.8 (25.5)		4.5 (22.7)		0.87 (0.53 - 1.36)
		Power	- 2.7 (23.8)		27.9 (27.0)		3.2 (28.3)		0.87 (0.54 - 1.38)
Bulb	Individual	Time	5.2 (1.3)	5.1 (1.5)	8.9 (3.2)	7.1 (1.3)	6.9 (1.3)	6.7 (1.1)	0.83 (0.64 - 1.18)
		Smoothness	- 7.42 (1.9))	- 7.9 (2.2)	- 10.5 (2.6)	- 9.5 (1.6)	- 9.0 (2.1)	- 9.2 (1.9)	0.62 (0.21 - 1.05)
		Speed	202.5 (54.4)	211.1 (49.6)	146.3 (51.9)	203.2 (53.7)	190.8 (56.2)	208.2 (48.8)	0.80 (0.55 - 1.17)
		Power	248.8 (82.0)	262.3 (91.0)	183.2 (79.8)	283.2 (104.5)	260.1 (106.8)	284.8 (92.6)	0.80 (0.57 - 1.16)
	Asymmetry	Time	3.0 (9.5)		25.9 (33.9)		2.9 (16.2)		0.85 (0.37 - 1.68)
		Smoothness	4.1 (17.9)		- 13.9 (42.1)		0.6 (22.2)		0.42 (- 0.08 - 0.79)
		Speed	3.4 (18.5)		28.8 (15.2)		8.8 (16.5)		1.23 (0.80 - 1.86)
		Power	2.6 (23.6)		35.6 (16.2)		9.9 (20.7)		1.35 (0.89 - 2.05)

Table 5.2. Demographic data and summary scores for matched control and patient groups at baseline and follow-up. ▒ columns represent non-dominant or affected shoulders, ▒ columns represent dominant or unaffected shoulders for control subjects and patients, respectively.

5.4.1 Reliability

Intra-rater reliability, based on unaffected shoulder scores for all patients (25 shoulders) at two time points, was demonstrated for all scores except task time. The highest intra-rater ICCs were seen with power scores (Back: 0.75 CI 0.51-0.88 Bulb: 0.84 CI 0.68-0.93). Inter-rater reliability, based on repeated assessment of ten control subjects (20 individual shoulder scores), was demonstrated on all scores. Inter-rater ICCs for smoothness showed moderate to good score reliability (Back: 0.82 CI 0.60-0.92, Bulb: 0.75 CI 0.47-0.90). Speed demonstrated moderate score reliability (Back: 0.63 CI 0.29-0.84, Bulb: 0.51 CI 0.11-0.77) and power score ICCs were considered poor to moderate (Back: 0.66 CI 0.33-0.85, Bulb: 0.48 CI 0.0.06-0.75). Task duration for both procedures demonstrated excellent reliability (Back: 0.91 CI 0.78-0.96, Bulb: 0.94 CI 0.84-0.97).

5.4.2 Known-group validity

Comparisons of baseline patient (pathological side) and control subject individual shoulder scores revealed significant differences for all scores, with pathological shoulders exhibiting reduced smoothness, reduced movement speed, reduced power and increased task time. For patient baseline assessments, all scores differed significantly between affected and unaffected shoulders except for movement smoothness. Similarly, differences were significant between both study groups for all asymmetry measures except smoothness (Figure 5.4). For patient assessments following treatment, only speed and power calculated during the bulb procedure differed significantly between affected and unaffected shoulders. No significant differences were found between patient follow-up asymmetry scores and control group asymmetry scores.

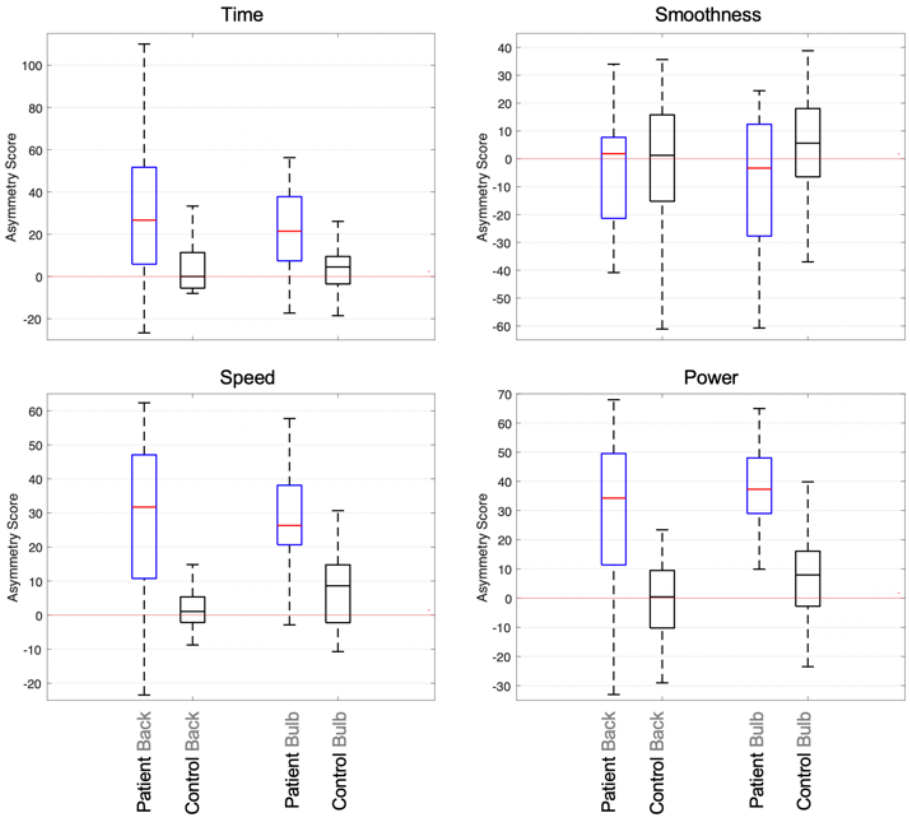


Figure 5.4. Box plots (centre bar: median; hinges: 1st & 3rd quartile; whiskers: 1.5 x interquartile range) for all asymmetry scores comparing patients (baseline scores) to control subjects for both *back* and *bulb* procedures.

ROC analysis found that all baseline individual shoulder scores were capable of differentiating groups, with time performing the best (Figure 5.5). Speed and power metrics demonstrated improved diagnostic power with asymmetry scores over individual shoulder scores (AUC 0.90 and 0.91 vs 0.85 and 0.86, respectively). Asymmetry scores for smoothness performed relatively poorly at differentiating groups (AUC 0.62).

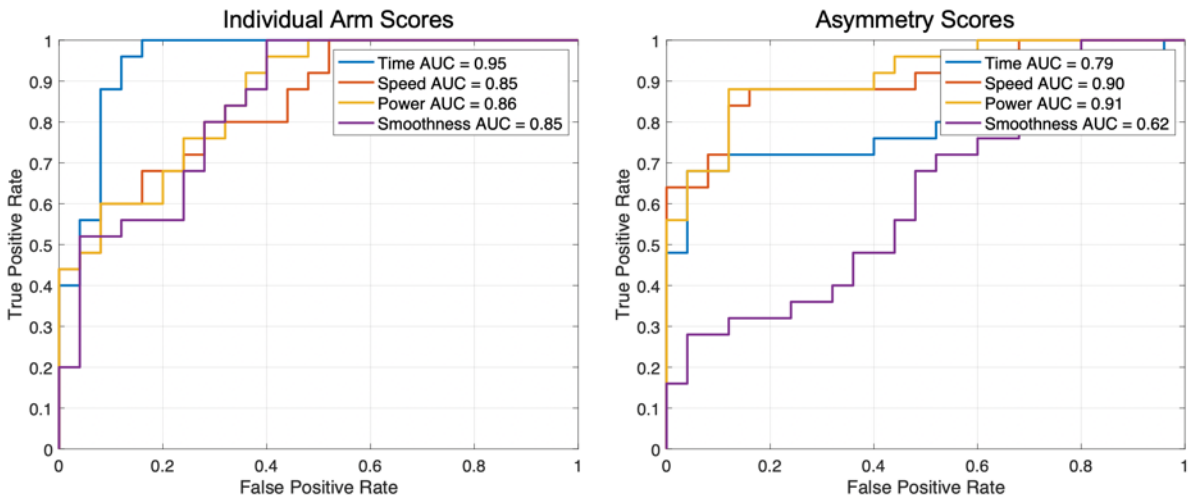


Figure 5.5. ROC curves with AUC value reported for individual shoulder scores and asymmetry scores discriminating between patients and healthy controls.

5.4.3 Score responsiveness

Changes in individual shoulder scores from baseline to follow-up were statistically significant, with affected shoulders demonstrating increased smoothness, speed and power following treatment (Figure 5.6). For asymmetry scores, all metrics except smoothness demonstrated significant changes, with effect sizes ranging from 0.42 to 1.35. Large effect sizes (> 0.8) were seen with all power and speed scores calculated during both components of the movement procedure. Time, speed and power scores exceeded the effect size seen with the OSS (0.70).

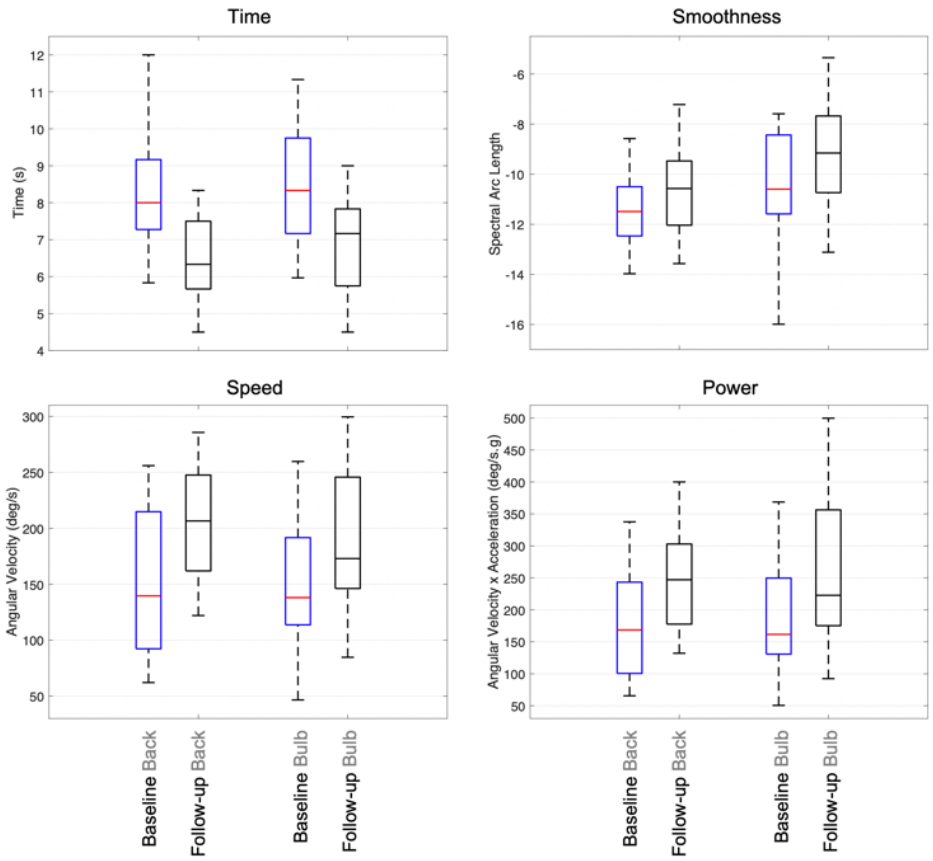


Figure 5.6. Box plots (centre bar: median; hinges: 1st & 3rd quartile; whiskers: 1.5 x interquartile range) for patient individual shoulder scores (affected side) comparing baseline to follow-up.

5.4.4 Concurrent validity

Both individual and asymmetry scores for speed and power had statistically significant correlation with the OSS (r = -0.40 – -0.58). Task duration during the back component of the procedure demonstrated a statistically significant but relatively weaker correlation (individual shoulder score r = - 0.32, asymmetry score r = - 0.34). Task duration during the bulb procedure and all smoothness scores demonstrated no statistically significant correlation with the PROM (Figure 5.7).

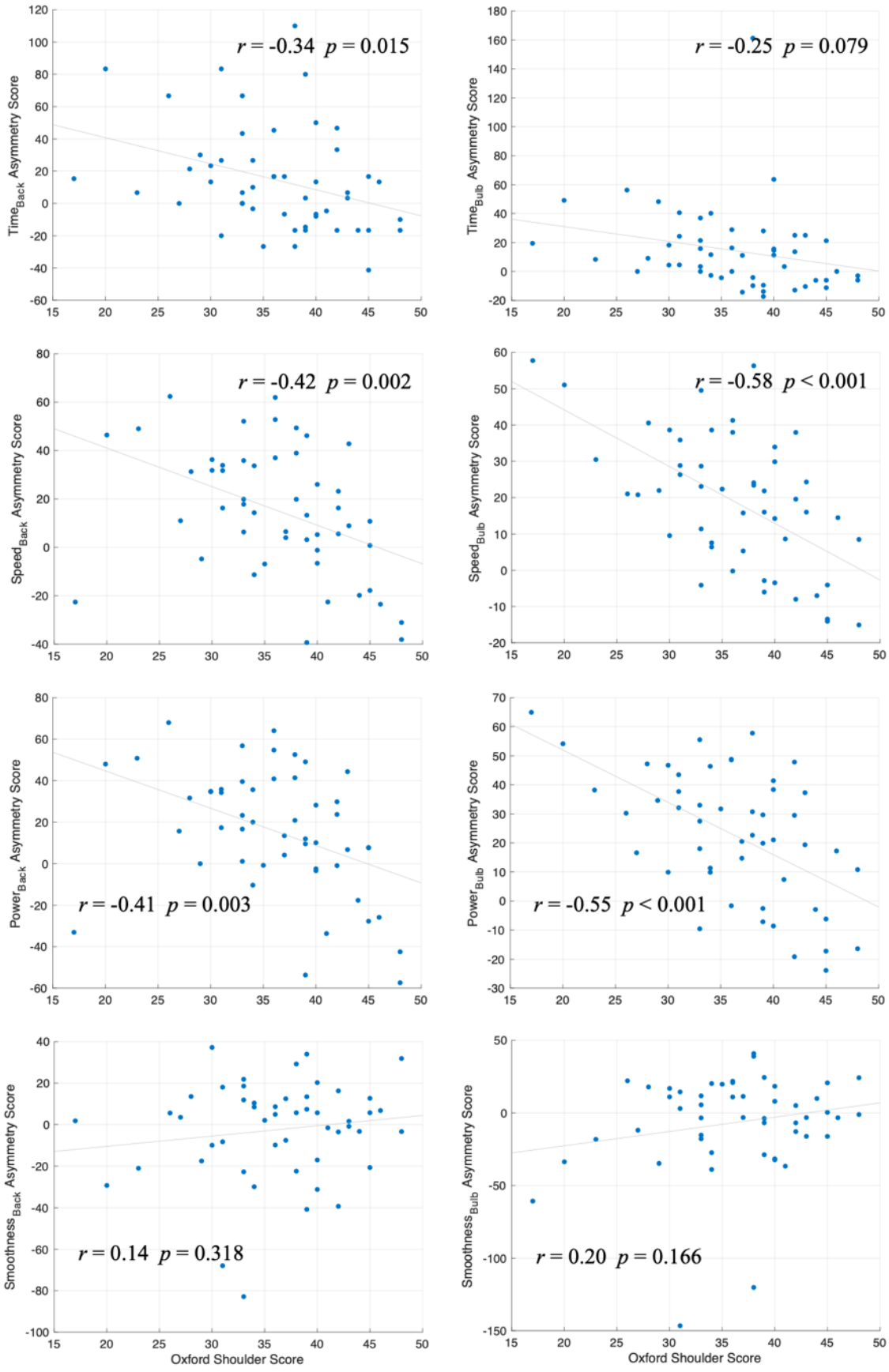


Figure 5.7. Scatter plots demonstrating the relationship between asymmetry scores and the Oxford Shoulder Score.

5.5 Discussion

5.5.1 Implications of study

Inertial measurement units are capable of measuring a range of human movement parameters, and can offer clinicians relevant insights into movement quality that could support the management of musculoskeletal conditions. This chapter explores the utility of inertial based scores in assessing shoulder function and includes an evaluation of the smoothness metric SPARC. Smoothness was found to display good psychometric qualities in evaluating patients with subacromial shoulder pain, demonstrating reliability and responsiveness to treatment. Consistent with previous studies, our investigation showed scores derived from accelerometer and gyroscope output ranges recorded during functional movement procedures to demonstrate excellent psychometric properties. As well as discriminant and convergent validity, these scores displayed better responsiveness to state changes than a well-established validated PROM.

Characterising limb function can be challenging due to the lack of consensus regarding its definition as well as subjectivity around what constitutes good function. The nature of any functional impairment may be specific to certain injuries and conditions, and the impact on one's ability to carry out daily living activities will be specific to certain populations. Peak movement speed and acceleration would be expected to be compromised in the context of pain; however, it has been suggested that these inertial scores are not related directly to pain but rather the pain's effect on mobility (68). The responsiveness of certain inertial scores relative to the OSS may support this, with greater improvement in inertial scores following treatment

potentially reflecting an improvement in clinical function not fully captured with conventional PROMs. Furthermore, correlation with the OSS demonstrates that the inertial scores are at least partially representative of patient-perceived function.

Smoothness scores have been shown to reliably identify affected limbs in patients with spastic hemiparesis, accurately assess gait patterns patients with Parkinson's disease and quantify functional recovery following stroke (74, 79, 80). Smoothness is a feature of well-developed motor behaviour, and the measurement of this quality naturally lends itself to the assessment of neurological conditions. While it is understood that factors at the level of the musculoskeletal system also contribute to this movement quality, little is known about smoothness profiles of patients with musculoskeletal pathology. The properties of the SPARC metric demonstrated in this study indicates that smoothness is also a clinically relevant functional domain in musculoskeletal conditions. Significant changes following treatment with moderate effect sizes show that this score has potential in assessment and monitoring of patients with subacromial pain. Further study is required to understand the influence of various factors on score profiles. Age was confirmed to be a significant score predictor in this study, but task familiarity and fatigue are also likely to be factors. As with speed and power metrics, a better understanding of movement smoothness is necessary for accurate clinical interpretation and optimal use of the scores.

The movement procedure implemented in this study incorporated rotation and elevation; two essential components of shoulder function, which are necessary for carrying out important activities of daily living (143). Perineal care requires internal rotation and general grooming tasks require elevation and external rotation (145). It is found that these movements are typically compromised in commonly occurring shoulder conditions (146), justifying their

use as kinematic scoring procedures in patients with subacromial shoulder pain. It is possible that the relatively unconstrained (i.e. no strict requirements on endpoint hand positions) nature of the movement procedure could introduce greater variability in the results. It is also possible that more complex movement patterns requiring greater levels of coordination or strength could be better suited at identifying relevant differences between pathological and healthy shoulder movement. This would certainly be the case for patients with higher functional demands beyond basic daily living activities, not reflected in the simple back to bulb procedure. Increasing task complexity could however exclude patients with more severe functional limitations, compromising the generalisability of the assessment.

Asymmetry scores for speed and power metrics performed well under validity testing, supporting this method of referencing affected arm movement parameters to that of the unaffected arm. This method has been used in similar studies (68-71, 143) and provides a convenient method of standardising scores in the context of what is normal movement for a particular individual. However, if patients are experiencing symptoms in their contralateral shoulder, the validity of the asymmetry scores would naturally be compromised. This study also found that individual arm scores for age matched control subjects were different to the patients' unaffected shoulder scores, in that they displayed increased smoothness, speed and power. The lower than expected score profiles of patients' asymptomatic arms could reflect a pre-clinical condition, as it is found that bilateral rotator cuff disease, either symptomatic or asymptomatic is common in patients presenting with unilateral symptoms (147). In any case, the findings suggest that an asymptomatic shoulder in patients with degenerative shoulder conditions may not be a reliable comparator.

5.5.2 Study limitations

The key limitation of this study is the relatively small sample of subjects assessed. Age was identified to be a significant predictor for score value, however it is possible that other variables such as BMI and handedness were not found to impact score value due to an insufficient sample for the analysis. Larger sample sizes would allow for a better understanding of these relationships and permit an analysis controlled for these variables.

Additionally, the patients evaluated were a heterogenous group with regards to the severity of their condition, baseline functional status and characteristics. It is likely that severity and chronicity of symptoms could influence movement quality and potential for recovery. Variation in treatment received could also influence score variability and compromise interpretation of results. Receiving treatment is no guarantee of improvement and successful recovery would be dependent on a number of variables, including treatment compliance. More controlled studies will need to be carried out to understand what score changes reflect clinically meaningful improvements.

5.5.3 Summary

Assessment using wearable devices is practical, affordable and most importantly, objective. Inertial sensor metrics demonstrate good psychometric properties and have the potential to complement existing methods of musculoskeletal assessment in clinical and research settings. The investigation in this chapter evaluates scores recorded during functionally relevant movements within controlled assessments. The value of these metrics in tracking patient progress are explored further in the Chapter 7 with automated serial

measurements taken during home rehabilitation activity. Before this could be achieved, processes to segment, recognise and classify shoulder activity from inertial data were required. This involved further development of the existing MATLAB code, and the creation of fully automated processing pathways that could be applied to interpret inertial sensor data in near real-time. These methods are described and validated in Chapter 6.

6 Rehabilitation Exercise Monitoring and Classification

Contents

6.1 Introduction 97
6.1.1 Adherence in musculoskeletal rehabilitation 97
6.1.2 Supervised machine learning for human activity recognition 98
6.2 Study aims 99
6.3 Methods 100
6.3.1 Study participants and data collection 100
6.3.2 Pre-processing 102
6.3.3 Segmentation 102
6.3.4 Feature extraction 105
6.3.5 Feature selection 105
6.3.6 Model training and validation 107
6.4 Results 114
6.4.1 Reliability of segmentation 114
6.4.2 Feature selection 115
6.4.3 Model training validation 117
6.5 Discussion 120
6.5.1 Implications of study 120
6.5.2 Study limitations 122
6.5.3 Summary 123

6.1 Introduction

6.1.1 Adherence in musculoskeletal rehabilitation

Exercise-based rehabilitation plays a central role in the management of subacromial shoulder pain and is often initiated as first-line treatment (10). Although most rehabilitation programs are guided by a specialist, patients are typically expected to perform the majority of their prescribed treatment independently in their home environment. It is well established that adherence to home musculoskeletal rehabilitation is poor, with studies reporting non-compliance to be in the region of 50-65% (148). With such variability and uncertainty in treatment implementation, it is almost impossible to draw meaningful conclusions about the efficacy of rehabilitation in both clinical and research settings. Existing self-reported measures of adherence perform poorly under psychometric testing and do not provide a detailed account of exercise activity (12, 40). There is, therefore, a clear need for robust methods to measure the implementation of exercise-based treatments (149).

Recent developments in inertial sensor technology have created opportunities for the objective monitoring and analysis of human activity (17). Quantifying rehabilitation activity in a meaningful way could prove to be a clinically useful indicator of patient progress and performance. Automated recognition of shoulder movement could also facilitate the calculation of kinematic parameters, such as those explored in Chapters 4 and 5, during relevant functional activity.

6.1.2 Supervised machine learning for human activity recognition

Supervised machine learning algorithms can be used to classify patient activity based on inertial sensor data (84), and it is felt that these methods could provide a practical and reliable approach to monitoring rehabilitation activity (86, 88, 150, 151). These techniques involve the use of labelled training data, containing samples of inertial data sets and their corresponding activity labels, to build predictive classification models capable of determining what exercises are being performed by the patient. Human activity recognition (HAR) systems could facilitate the evaluation of treatment fidelity within clinical and research settings and have the potential to provide patients with meaningful feedback (15, 152). Qualitative studies also indicate that this form of feedback could drive patient motivation and help maintain or improve adherence to home exercise programs (153).

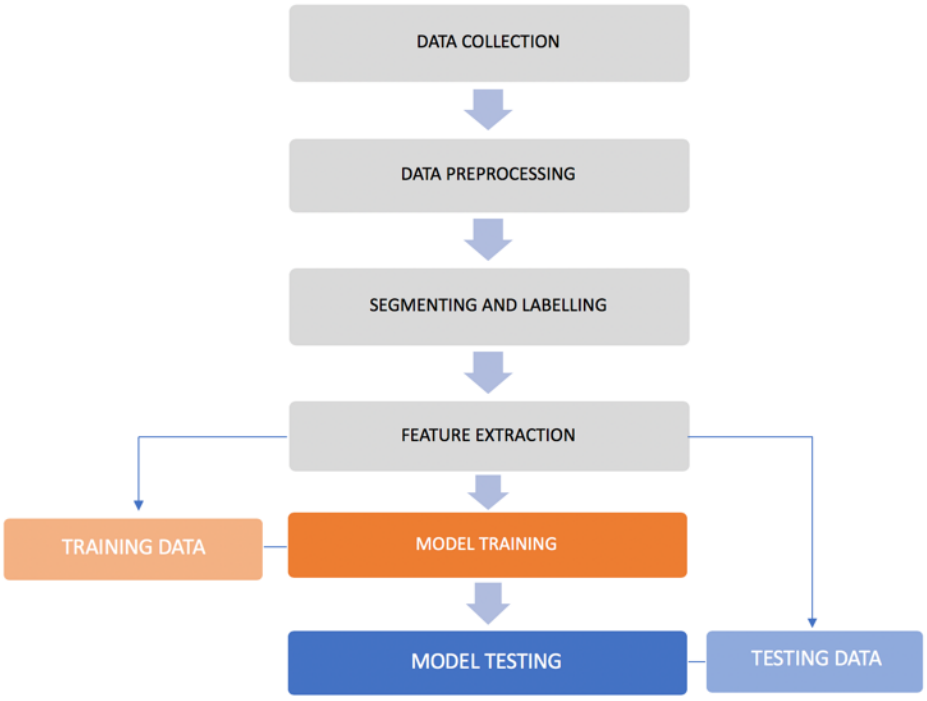


Figure 6.1. Overview of activity classification study implementing supervised machine learning techniques.

Results from recently published work appear promising; however, most classification systems have been validated with inertial data acquired from healthy study participants (15). Patients with pathology will naturally display greater variation in movement pace and trajectory, and trained classifiers will need to accommodate for this variation. Additionally, existing studies do not evaluate classifiers' ability to differentiate target activity from non-specific or irrelevant movements typically present within inertial recordings. This process has previously been achieved by manual or semi-automated extraction of target activity data, which increases user burden and would not be compatible with real-time monitoring applications.

6.2 Study aims

The aim of this study was to evaluate the performance of a fully automated classification pathway for recognition of shoulder rehabilitation activity carried out by patients with subacromial shoulder pain. Supervised machine learning models were used to classify shoulder activity from inertial data acquired through a single wireless MIMU. Establishing a reliable method to segment data streams and calculate informative and discriminating features, were the main technical challenges to delivering this study.

6.3 Methods

6.3.1 Study participants and data collection

This study was granted ethical approval by a UK NHS Research Ethics Committee (Appendix B). Patients with a diagnosis of subacromial shoulder pain and who were deemed suitable for outpatient physiotherapy were recruited from NHS orthopaedic clinics. Criteria for exclusion included adhesive capsulitis, inflammatory arthropathy, any neuromuscular condition, cervical spondylosis or a history of recent trauma. Twenty consecutive patients (eight males, twelve females) meeting the eligibility criteria provided informed consent for their participation in this study. The recruitment target was chosen to ensure collection of sufficient movement data to represent that of the target population, without over-recruiting. Five patients had partial or full degenerative rotator cuff tears, and 18 patients had previously received some form of treatment in the form of physiotherapy or steroid injections. Mean participant age was 58.7 years (SD 9.4), and 18 participants were right arm dominant. Assessments were performed on the affected arm only (ten left ten right), and the mean Oxford Shoulder Score was 30.9 (SD 8.3).

Participants were asked to review a patient information booklet detailing five rehabilitation exercises routinely prescribed for the treatment of subacromial shoulder pain: shoulder abduction, shoulder flexion, wall slide, wall press and shoulder rotation (Figure 6.2). Wall slide, wall press and rotations were chosen as these exercises form the British Shoulder and Elbow Society recommended home rehabilitation program for subacromial pain (154). Abductions and flexions were included to evaluate the classifiers' ability to differentiate similar

single movements carried out in different planes. Under supervision, participants were instructed to carry out ten repetitions of each exercise at a self-selected pace and advised to stop if they experienced significant discomfort. All exercises were carried out with the patient standing, except the shoulder rotations, which were carried out from a seated position with the elbow supported to keep the upper arm abducted close to 90°.

Raw nine-axis (three-axis accelerometer, gyroscope and magnetometer) motion data was acquired wirelessly via a single MIMU mounted to an arm sleeve worn above the level of the elbow. Accelerometer, gyroscope and magnetometer sampling rates were set to 100Hz, 100Hz and 25Hz, respectfully. Data was collected wirelessly via Bluetooth Low Energy to a smart device and processing was performed offline in MATLAB (R2018a from MathWorks).

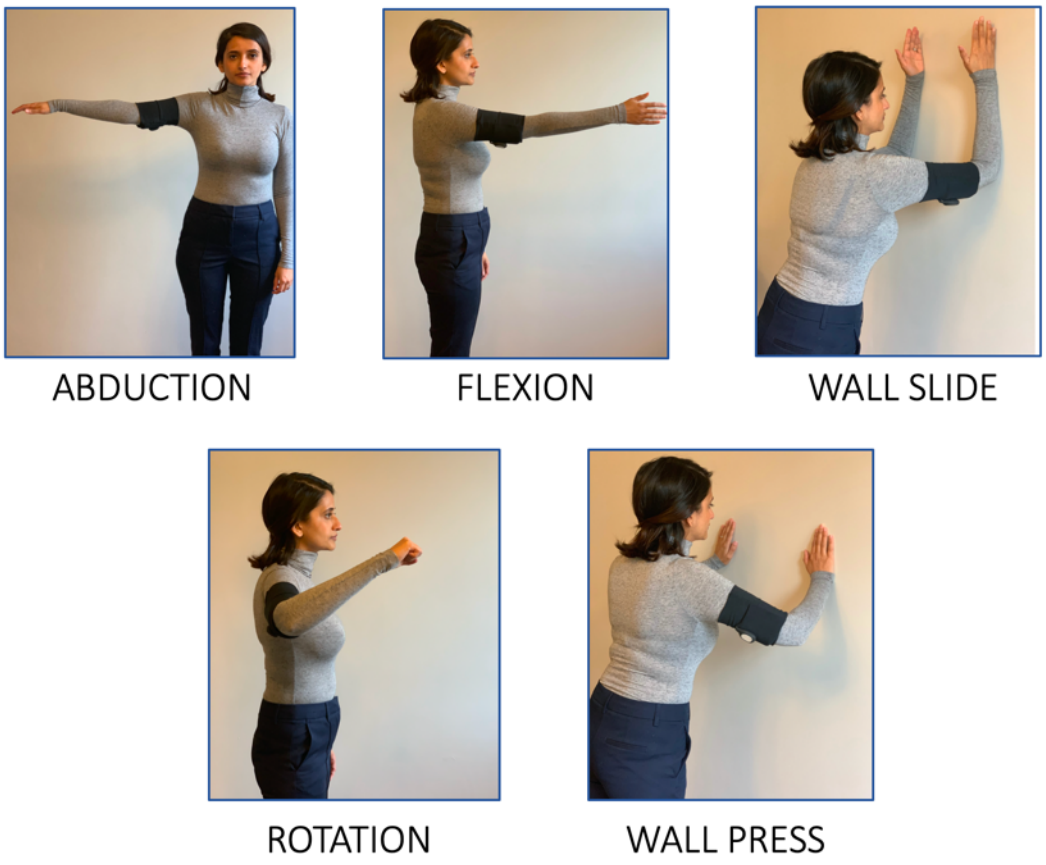


Figure 6.2. Ten repetitions of five exercises recorded with a single arm mounted MIMU.

6.3.2 Pre-processing

Pre-processing was performed to create a consistently sampled data set suitable for feature extraction. Magnetometer data was up-sampled to 100Hz, and a low pass sixth-order Butterworth filter (10Hz cut-off frequency) was applied to the accelerometer data to remove high-frequency noise. Further detail on pre-processing can be found in Chapter 3. In addition to the individual pre-processed nine-axis data streams, features were also calculated from individual sensor vector magnitude values and three-dimensional estimates of sensor orientation. The Madgwick orientation filter was used to fuse nine-axis sensor data to calculate sensor orientation (103) and features were calculated from both quaternion and Euler angle (*ZYX* sequence) representations of arm orientation.

6.3.3 Segmentation

Segmentation of pre-processed data streams is required to produce packets of data to represent specific activity. This is performed using windowing techniques, which consists of dividing data streams into smaller time segments. Classification algorithms are subsequently applied to data within each window to classify activity. The standard windowing method involves dividing data streams into fixed-length segments, and to improve the efficiency of this technique, windows often have some degree of overlap (84). This technique lends itself well to most HAR applications, for example, when the objective is to identify when a patient is sitting, standing or walking. However, for rehabilitation activity, we would ideally need each window to be of variable length to dynamically represent a specific activity cycle, irrespective of the time taken to perform the activity. This would facilitate more reliable classification and importantly allow one to quantify the amount of exercise performed.

For this study, I developed and implemented a novel event-defined windowing technique. This method involved identifying the dominant gyroscope axis zero-velocity time points to define the boundaries of each movement cycle. The dominant axis naturally varies depending on movement trajectory; therefore, a peak analysis function was applied to the gyroscope vector magnitude data stream to identify time points when the velocity was closest to zero (Figure 6.3). Parameters of the peak analysis function that needed defining included minimum peak height and minimum peak distance. The MATLAB code written to achieve this is provided in Appendix D. Unlike conventional fixed-size sliding windowing techniques, this method potentially ensures each segment represents a defined portion of an activity. This stage was fully automated, and all recorded data segments (exercise and non-exercise activity) were included for analysis.

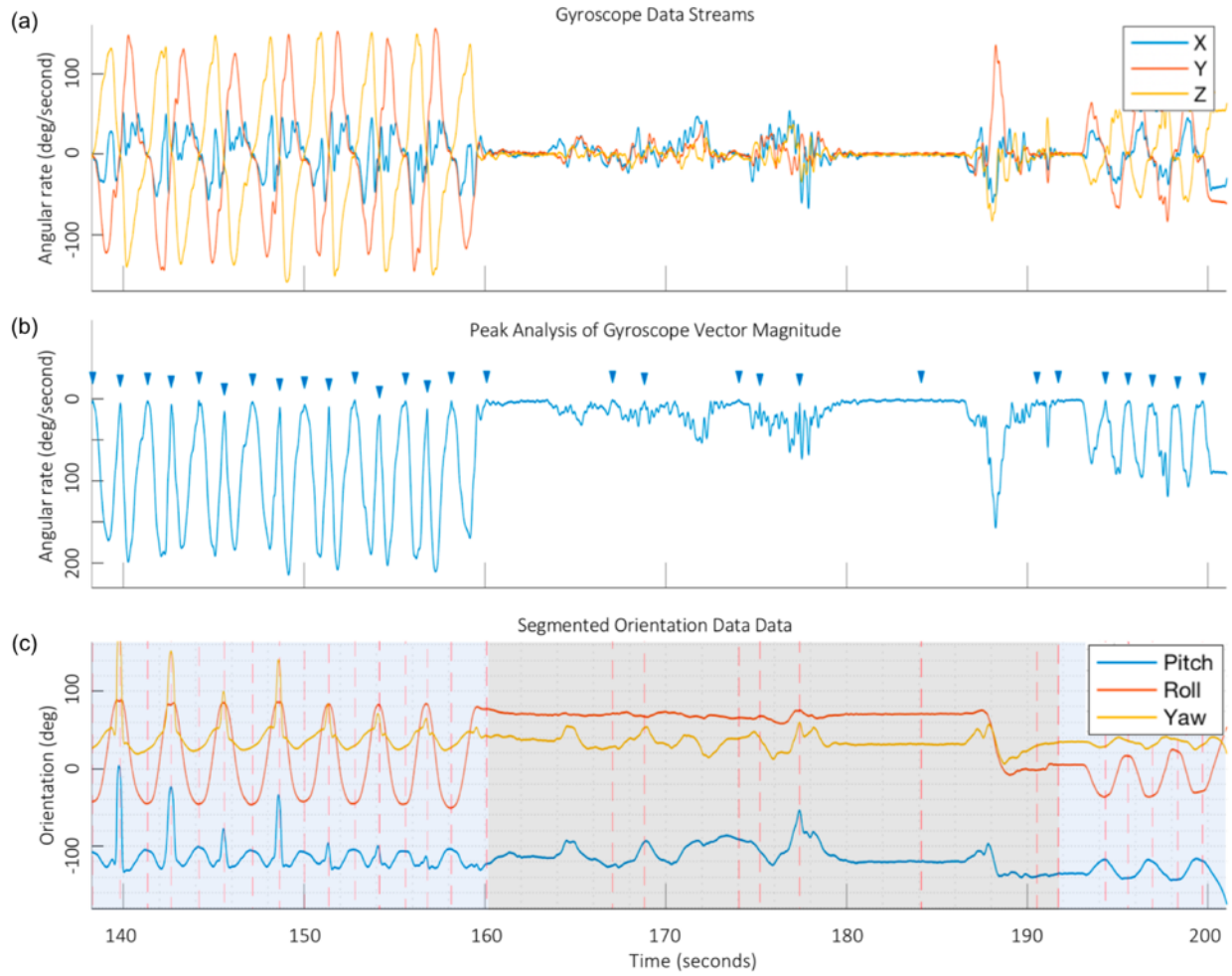


Figure 6.3. (a) 3-axis gyroscope plot; (b) Peak analysis of (inverted) gyroscope vector magnitude identifies locations close to 0; (c) Locations of identified peaks superimposed on orientation data indicates segmentation points.

A unique exercise label was attributed to each activity segment from ground truth data collected during performance of the exercises. Two adjacent segments defined a single exercise due to the zero-velocity state at maximum shoulder excursion (e.g. peak of abduction). Both segmented components of a single exercise (e.g. abduction and adduction) were assigned the same label (*Abduction*) providing five activity classes. A sixth class (*Junk*) represented non-specific activity recorded during the session, which was unrelated to the execution of the prescribed exercises. This included inertial data recorded while patients adjusted body or arm

position in between exercises, adjusted clothing or performed upper limb gestures while communicating. Accuracy of ground truth data sets was verified with offline visualisation of orientation data streams, which provided an interpretable representation of arm position and movement, allowing the investigator to infer exercise type and confirm label.

6.3.4 Feature extraction

The feature extraction process calculates the unique and distinguishing characteristics of each activity segment, which are used as input for the classification algorithms. This study evaluated a wide range of time and frequency domain features derived from accelerometer, gyroscope, magnetometer and orientation data (155). Time-domain features included mean, root mean square, standard deviation, variance, range, inter-quartile range, percentiles (5th, 25th, 50th, 75th, 95th) and vector pair (*xy/xz/yz*) Pearson correlation coefficients (bivariate features). Frequency domain features included maximum frequency component, mean frequency component, energy spectral density, entropy and kurtosis. A total of 237 features were calculated. These calculations were automated and carried out for each activity segment within a for-loop (programming control flow statement for specifying iteration), using formerly determined segmentation points to determine which portions of the data stream each round of calculations are performed on.

6.3.5 Feature selection

Feature selection aims to reduce the number of features used by identifying those that are of most independent predictive importance (156). This step optimises the learning process and improves classification accuracy by reducing the risk of model overfitting (84). Optimal

feature selection can be achieved through three methods. *Wrapper* methods involve training and testing a specific algorithm using various or even all possible combinations of features (157). Given this method uses a classifier in the selection process, it typically yields optimal results; however, it is extremely computationally demanding and has to be performed for each classifier intended for use. *Filter* methods apply directly to the data set by reviewing the intrinsic properties of the features independent of what classification algorithm is being used. This allows features to be ranked based on independent value, but may not evaluate performance of specific features along-side others, failing to reveal redundancies in the feature set. This method requires less computation than wrapper approaches and provides a general feature selection strategy for most learning algorithms. Finally, *hybrid* methods select the most relevant features based on the internal parameters of a specific classification algorithm, and may not be applicable to all classifiers (84).

Given a large number of features within this study and the variety of algorithms being investigated, an established filter method was implemented. The ReliefF algorithm was applied to assign weight values to features, penalizing those that give similar values to segments of a different class and rewarding those that give different values to segments of different classes (158). The ReliefF algorithm exists as a built in MATLAB function, and was applied to the entire feature set for all activity segments from all patients. A subset of features was selected after setting an empirical threshold based on the calculated average of attributed weights, all features below the threshold were excluded. All features included for classifier input were scaled to zero-mean and norm-standard deviation.

6.3.6 Model training and validation

To solve a classification problem, two broad categories of learning algorithms can be considered. *Supervised* machine learning involves building a model using manually labelled data. The model can then be applied to identify class labels for new observations on the basis of training data with known class labels. *Unsupervised* learning algorithms do not require labelling of data and instead work independently to discover patterns or information held within the data. This naturally requires less manual intervention and can be used to identify patterns that may not be apparent to the investigator. Successful application of unsupervised machine learning techniques can be found in the fields of genetic analysis (159) and these methods are also commonly used in market research; however, HAR studies have demonstrated unreliable results using these techniques (84). As a result, this study implements supervised techniques, and four traditional machine learning classification models were chosen for evaluation:

***k*-nearest neighbour (*k*-NN)** algorithms categorise new input data based on the classes of their nearest neighbours in the training data feature space. K-NN predictions assume that features in the same proximity are from a similar class and various distance metrics can be used to find the nearest neighbour. This is a simple algorithm that can establish benchmark learning rules, but requires storage of all available training data, and therefore memory requirement is high.

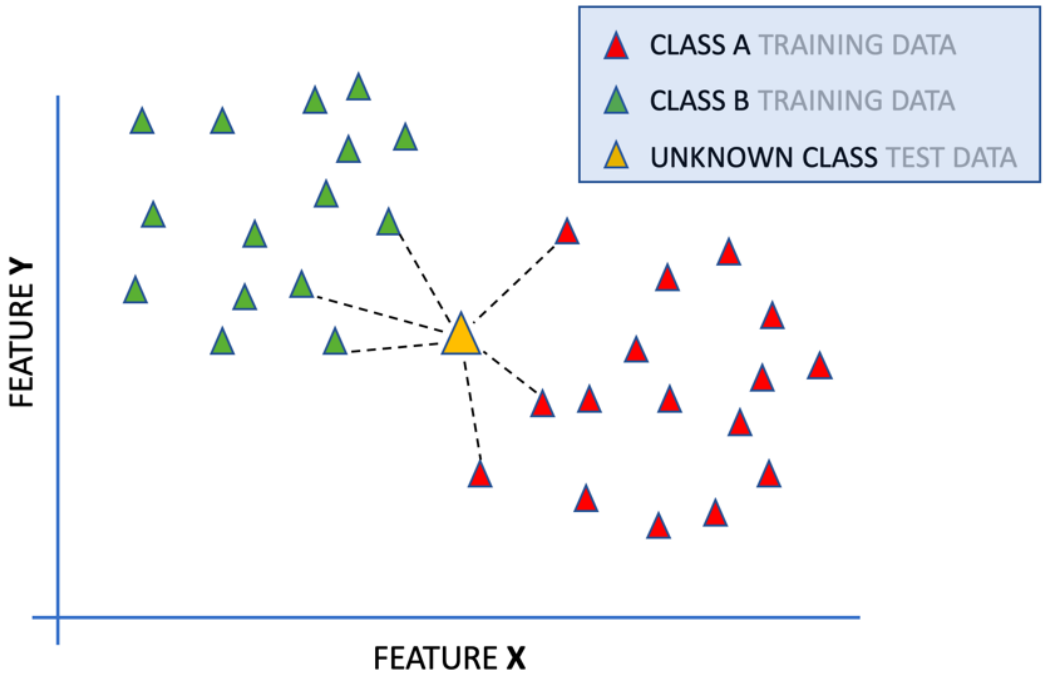


Figure 6.4. Feature mapping of *k*-nearest neighbour algorithm with dashed lines representing distances calculated between new class data and neighbouring training data. Relative distances between neighbouring training data classes will determine class allocation for test data.

Decision trees (DT) are commonly used algorithms that can applied to solve both regression and classification problems. This algorithm uses a tree representation to solve a problem where each node represents a feature, each link represents a decision, and each leaf represents an outcome. The value of a predictor is compared to a trained weight, and the number of branches and the value of the weights are determined in the training process. This closely mimics the human decision-making process and is, therefore, easy to understand and interpret. Other advantages include minimal processing requirements and fast prediction speed. However, decisions trees do not tend to provide optimal prediction accuracy and tend to overfit.

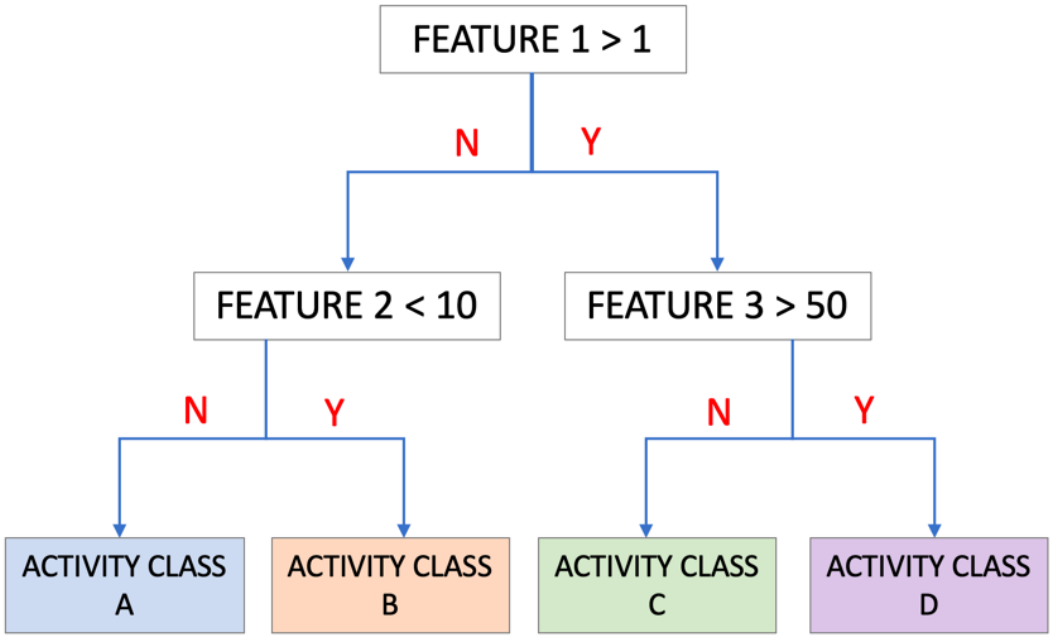


Figure 6.5. Simple decision tree.

Random forests (RF) are ensemble classifiers, in this case combining several weaker decision trees to create a stronger ensemble. This particular classifier represents a bagged DT, which contains multiple trees trained independently on data that is bootstrapped from the original data set. As predictors are randomly selected for each tree, there is a low correlation between the individual models, and this is key to the high accuracy of this classifier. Unlike simple DTs, RFs are less likely to overfit, but naturally require greater computational resources and have longer prediction times.

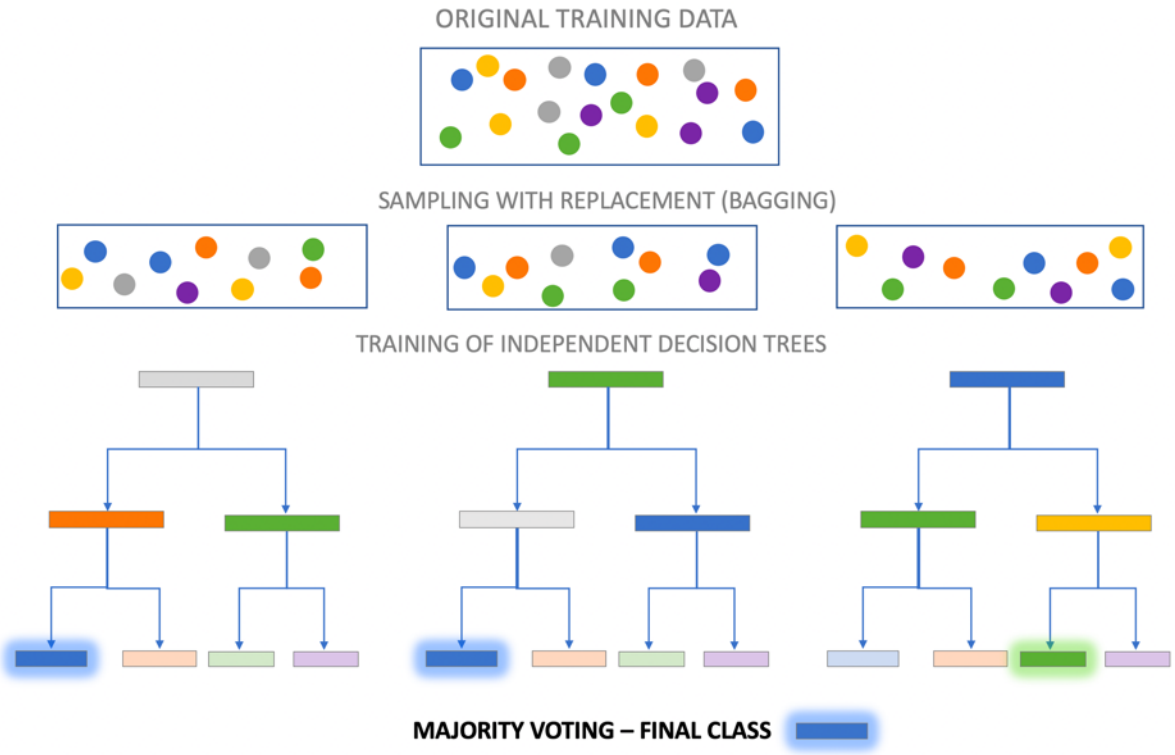


Figure 6.6. Random forests algorithm with individual trees trained on data bootstrapped from original data set.

Support vector machine (SVM) classifies data by establishing a linear decision boundary (hyperplane) that separates feature values of one class from feature values of another class. This requires mapping the original training data set and the optimal boundary is one with the largest margin between different classes. In some cases, non-linearly separable data can be transformed into a higher dimension where a linear decision boundary can be found (84).



Figure 6.7. Support vector machine with hyperplane created within largest boundary found between two classes.

All four algorithms were trained using the same selected subset of high-ranking features and were tested in a similar way. *Accuracy* was used to provide an overall indicator of model performance and reflects the proportion of correctly classified segments within testing. The classic *holdout* method to evaluate classifier accuracy involves dividing the available data into two parts: a training and a testing set. The model is created using the training set data and the performance is then evaluated with the test set. The main limitation of this validation method is potential for classifier performance to be affected by the composition of training and testing sets. A relatively small training set will result in model bias, as the model is not trained on a sufficiently representative volume of data. A relatively small testing set, however, could result in high variance in model performance. *Cross-validation* helps resolve this problem by using all data for training and testing and aims to achieve an appropriate bias-variance trade-off. Two distinct cross-validation methods were implemented in this study:

> ***k*-fold cross-validation** is a routinely applied method to estimate how well a predictive model will perform in practice (160). The entire labelled data set is divided into k number of folds, with an equal distribution of activity classes. One round of cross-validation involves model training with $k - 1$ folds, followed by testing with data from the omitted fold. k number of rounds are performed using different partitions, and validation results are combined to give an estimate of model performance. This method allows the use of all data for training and testing and protects against model overfitting. For this study k is given a value of ten, which is a commonly used value in practical machine learning scenarios to achieve an optimal bias-variance trade-off.

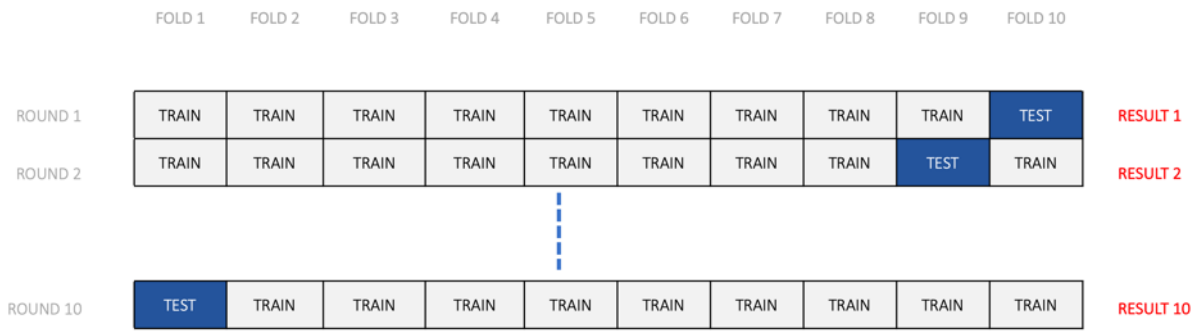


Figure 6.8. *k*-fold cross-validation, $k = 10$.

Leave-one-subject-out-validation (LOSOV) partitions the data set into subject-specific folds with each fold containing only the data from a single participant. Similar to the previous method, a single fold is retained each round for testing, and a single performance estimate is reported as the average of all testing results. As with ten-fold cross-validation (10FCV), the models are always tested on unseen data. However, in this case the models being tested do not have the advantage of having been trained on similar data from the same subjects.



Figure 6.9. Leave-one-subject-out cross-validation.

For both 10FCV and LOSOV, the overall accuracy metric was defined as follows:

$$\text{Accuracy} = \frac{\text{Number of accurately predicted segments}}{\text{Total number of predicted segments}}$$

The accuracy metric does not take unbalanced datasets into account and can be biased to favour the majority classes. Therefore, individual label *sensitivity* (true positive rate or recall), *positive predictive value* (PPV; precision) and *specificity* (1-false positive rate) were also calculated. Algorithm training and 10FCV was performed using the MATLAB Statistics and Machine Learning Toolbox. Automated LOSOV was not an available function within the toolbox and therefore had to be performed manually. Once models were trained with data from 19 patients, the model code was exported and applied to make predictions on the remaining patient's data. This process was repeated for every iteration/patient.

6.4 Results

The segmentation protocol yielded 2,568 samples. Segments representing exercise activity had a mean duration of 1.78 seconds. Not all patients were able to perform ten repetitions of all five exercises due to movement restrictions related to their shoulder pain. Any additional repetitions and limited attempts at exercise performance recorded within the session were labelled and included in the analysis.

6.4.1 Reliability of segmentation

Two sources of forced segmenting errors were identified during labelling. *Over-segmenting* resulted in single activity cycles being divided into two or more segments and occurred in 2.7% of activity segments. *Under-segmenting* resulted in a registered segment incorporating more than one half of an exercise cycle and occurred in 1.8% of activity

segments. In both cases, segments were assigned labels according to the activity being performed. Exercise activity labels were fairly evenly distributed; 344, 355, 348, 314 and 418 segments represented abduction, flexion, wall press, wall slide and rotation activity, respectively. Junk activity formed the most abundant class with 789 segments, representing 31% of all data segments.

6.4.2 Feature selection

The feature selection strategy resulted in the top 120 features being included; these were predominantly time-domain features derived from the accelerometer, magnetometer and processed orientation data. Feature ranking identified bivariate features of all data streams as offering the highest discriminative value for exercise classification. Only a small proportion of frequency domain and gyroscope derived features were weighted sufficiently for inclusion (Figure 6.10). A full list of features calculated and their respective weights are detailed in Appendix C.

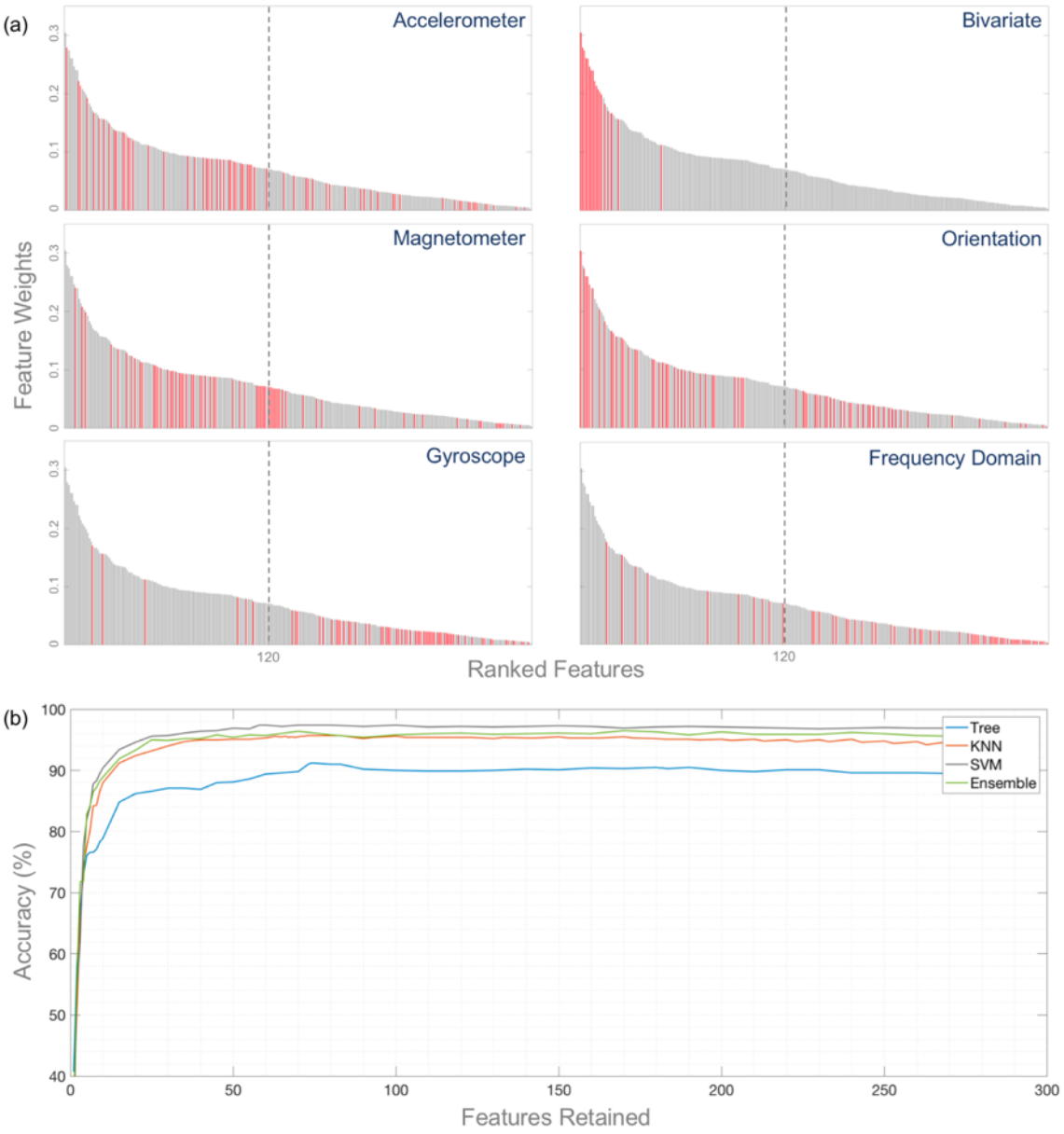


Figure 6.10. (a) All features ranked according to ReliefF algorithm weights (high to low) with individual plots highlighting specific feature subsets. (b) Relationship between feature inclusion (in order of ranking) and classifier performance (10FCV).

6.4.3 Model training validation

Performance metrics for each model with both validation methods are detailed in Table 6.1. The 10FCV results indicate how the system would perform if a subject-specific training set was created for each new patient. For this validation method, models achieved accuracies between 90.9% (DT) and 97.2% (SVM). SVM sensitivity scores, measuring the effectiveness of the classifier to identify a specific activity, exceeded 96% for all classes. PPVs were equally high, and all specificity scores exceeded 99%.

The LOSOV evaluates the system's ability to generalise across to entirely new patient activity. For all classifiers, the LOSOV method yielded lower overall accuracy scores and higher variation than 10FCV. DT classifiers were particularly weak at generalising to new subject data, with specific class sensitivity scores as low as 50.9% (abduction). The RF classifier performed significantly better than all other models with 80.5% overall accuracy. There was considerable variation in model performance depending on the activity being classified; sensitivity scores for wall slides and abductions were 93.4% and 62% respectively. LOSOV results also showed significant variation in individual subject fold accuracies.

	10-Fold Cross-Validation (%)				Leave-One-Subject-Out-Validation (%)			
	Accuracy	Sensitivity	PPV	Specificity	Accuracy	Sensitivity	PPV	Specificity
DT	**90.9** (±2.0)				**71.0** (±16.9)			
Abduction		86.9	86.4	97.9		50.9	62.9	95.4
Flexion		89.9	87.2	97.9		51.8	63.2	95.2
Wall Slide		96.3	95.4	99.3		82.5	83.2	97.4
Wall Press		93.6	92.2	98.9		67.2	75.6	97.0
Rotation		93.8	94.5	98.9		74.6	87.6	97.5
Junk		88.2	90.3	95.8		87.7	67.9	81.6
K-NN	**95.7** (±1.2)				**76.1** (±13.5)			
Abduction		97.4	94.4	99.1		63.7	67.4	95.2
Flexion		96.9	94.2	99.1		61.4	71.0	96.0
Wall Slide		97.1	97.7	99.6		85.3	81.1	96.9
Wall Press		98.1	93.9	99.1		74.5	82.7	97.8
Rotation		97.4	95.1	99.0		74.2	92.3	98.8
Junk		91.9	97.2	98.8		88.7	73.6	85.9
SVM	**97.2** (±1.2)				**74.9** (±17.0)			
Abduction		96.8	96.0	99.4		56.4	68.8	96.0
Flexion		96.6	96.9	99.5		55.5	75.2	97.1
Wall Slide		97.1	98.0	99.7		89.1	83.8	97.3
Wall Press		98.4	98.4	99.8		58.0	96.8	99.7
Rotation		97.8	97.4	99.5		74.9	97.5	99.6
Junk		96.8	97.0	98.7		95.7	65.9	78.1
RF	**96.4** (±1.7)				**80.5** (±15.6)			
Abduction		95.9	94.6	99.1		62.2	76.4	95.2
Flexion		94.6	94.4	99.1		67.6	75.9	96.0
Wall Slide		96.6	98.0	99.7		93.4	99.1	96.9
Wall Press		97.8	97.5	99.6		65.3	94.0	97.8
Rotation		98.6	97.2	99.4		90.0	95.4	98.8
Junk		95.7	96.7	98.5		94.6	72.3	85.9

PPV: positive predictive value; DT: decision tree; K-NN: k-nearest neighbour; SVM: support vector machine, RF: Random Forests.

Table 6.1. Performance of all four classification models using 10FCV and LOSOV.

The proportions of correctly and incorrectly predicted activity labels can be reviewed within a confusion matrix (Figure 6.11). This demonstrates the tendency for abduction, flexion and wall press activity to be misclassified as junk activity, and for abduction and flexion activity to be confused for one another.

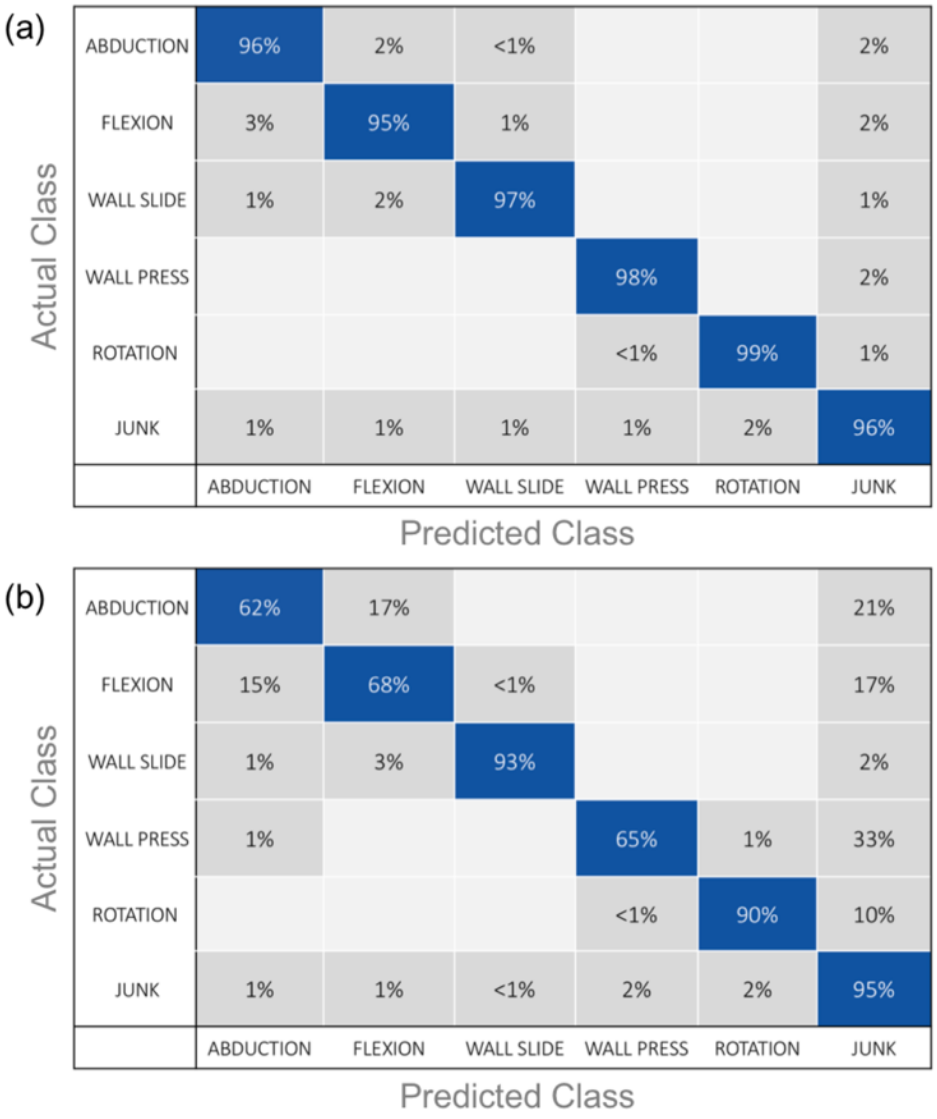


Figure 6.11. Confusion matrix of (a) 10FCV and (b) LOSOV validation results of single model (Random Forests).

6.5 Discussion

6.5.1 Implications of study

This study reports on the performance of supervised machine learning techniques in classifying shoulder rehabilitation exercises from inertial data acquired from a single nine-axis MIMU device worn on the upper arm. Classifier performance was evaluated with the use of real patient data and the inclusion of non-specific activity, thus providing representative estimates of their capabilities in real-world scenarios. The predictive performance demonstrated suggests that these systems can provide a clinically useful account of exercise adherence.

The classifiers chosen for this study represent a range of supervised machine learning models successfully implemented in previous HAR studies. The choice of classifier for future clinical platforms will rely not only on predictive accuracy but also on factors related to training time, prediction speed and memory usage. The SVM performed well for both validation methods and are highly regarded for their ability to handle high-dimensional data. Choosing appropriate model parameters for SVM classifiers can be complicated however, and like *k*-NN classifiers, they can be memory intensive. Simple DTs are quick to train and easy to interpret but are typically less accurate as this study has demonstrated. This RF classifier performed well for both validation methods and has shown excellent predictive accuracy in previous studies, outperforming SVM and *k*-NN algorithms (88, 161). The improved predictive capabilities and relatively fast prediction speeds make this algorithm a promising candidate for sensor-based rehabilitation monitoring (162). Deep learning techniques, utilising neural network-based algorithms, provide an alternative method for activity classification (88, 150). A key advantage

of these methods is their ability to learn directly from sensor output data, removing the need for manual feature engineering. However, the models do require substantially longer training times compared to traditional supervised classification methods and typically demand larger volumes of training data.

Accurate classification of human activity from sensor data is reliant on appropriate segmentation methods. The employed technique identified gyroscope zero-velocity crossing locations to define activity boundaries. This alternative method to conventional fixed-size overlapping windows was chosen for this study as establishing an optimal fixed window length can be challenging, and presents a trade-off between algorithm responsiveness and model performance (163). Furthermore, short and long-term factors such as disease progression, recovery, pain, fatigue and stress can cause significant variation in the time taken to carry out certain exercises. Segmenting based on gyroscope zero-velocity crossings ensures the windows capture a defined activity cycle and system accuracy will not be directly related to inter-subject and intra-subject variability in exercise performance time. This study demonstrates that zero-velocity segmentation is reliable and consistent for patient rehabilitation monitoring, and can therefore facilitate calculation of the performance metrics explored in Chapter 4 and 5, such as RoM, power, speed and smoothness. Additionally, this method provides a direct measurement of activity duration and the number of repetitions performed; both valuable metrics in the assessment of exercise adherence and performance (56).

Feature sets explored in this study were selected based on previous investigations evaluating HAC with inertial sensor data. The feature selection strategy employed an established *filter* method to evaluate the intrinsic properties of the calculated features (164). Our investigation supports previous findings on the relatively higher value of accelerometer

over gyroscope derived features, and time-domain over frequency domain features for these applications (165). Features derived from the magnetometer and processed orientation data (9-axis sensor fusion) were also shown to have high predictive value. Although, it must be noted that magnetometer readings are susceptible to distortions caused by ferromagnetic materials and the testing environment could influence the reliability of these features (47). Bivariate features of all data streams were shown to have the highest predictive importance; however, a recent study evaluating similar activities in healthy subjects found correlation coefficients of vector pair combinations to have poor predictive value (88). This discrepancy demonstrates the influence of data collection methods and windowing techniques on the differentiating qualities of calculated features (166).

6.5.2 Study limitations

The exercises chosen for analysis only incorporate single plane movements. Whilst these exercises are derived from recognised strategies for the treatment of subacromial pain, some multi-planar movement exercises can also be prescribed which have not been evaluated in this study. Data were collected from a relatively small group of patients and improved performance is likely to be achieved with larger amounts of training data. (86). The relatively poor differentiation between abduction and flexion movements within this study could also potentially be improved with an additional body sensor or a functional calibration procedure. Additional sensors could enable measurement of posture, scapulohumeral rhythm and movement trajectory, allowing an assessment of exercise performance (59). This would, however, increase patient burden, and future work should be aimed at understanding the trade-off between system performance and usability (15).

The activity was recorded during a single observed session for each subject, and it is uncertain how well the trained algorithms would generalise to subsequent independently performed exercise sessions where recorded movement patterns may change. A disadvantage to supervised learning methods is the necessity for labelled training data. Labelling is an expensive process, and there are significant challenges around collecting labelled data from home environments. Unsupervised learning systems have been implemented for HAR studies; however, predictions are less accurate and interpreting the output can be challenging (84). Semi-supervised learning approaches, which attempt to utilise large volumes of data with reduced labelling requirements, may offer the best balance between process efficiency and system accuracy (161).

6.5.3 Summary

Demonstrating adherence and program differentiation is critical in the development and evaluation of evidence-based rehabilitation interventions. This study demonstrates the feasibility of patient monitoring and automated classification of single plane exercises with wireless sensor technology. Chapter 7 will focus on evaluating the technical performance of these systems for home rehabilitation monitoring and understanding factors around usability.

7 Evaluation of Sensor-Supported Rehabilitation

Contents

7.1 Introduction 125
7.1.1 Connected health solutions for musculoskeletal rehabilitation 125
7.1.2 A prototype system design 125
7.1.3 System output 128
7.2. Study aims 129
7.3. Methods 129
7.3.1 Processing code development 129
7.3.2 Study participants and setup 130
7.3.3 Evaluation of activity recognition 132
7.3.4 Performance scores 132
7.3.5 Usability and perceived impact 133
7.4 Results 134
7.4.1 Accuracy of activity recognition 134
7.4.2 Performance score profiles 136
7.4.3 Usability 138
7.4.4 Perceived impact 141
7.4.5 Improvements 143
7.5 Discussion 144
7.5.1 Implications of the study 144
7.5.2 Summary 146

7.1 Introduction

7.1.1 Connected health solutions for musculoskeletal rehabilitation

Accessibility and reducing costs of mobile smart devices, wearable sensing technology and cloud computing is providing new opportunities to monitor and assess patients undergoing musculoskeletal rehabilitation. Connected health solutions can improve our understanding of the rehabilitation process, help develop effective treatments and generally enhance the patient experience. It is found that usability is one of the main barriers to widespread uptake of connected health applications (167). Usability should therefore be a priority when developing this form of technology, and should be evaluated alongside the technical capabilities of any wearable platform. To maximise on the potential value of biofeedback systems, their design should adopt a user-centred iterative process (168), where developers engage with end-users to understand the perceived benefits and obstacles around implementation in real-world scenarios (169). There is currently a distinct lack of both technical and usability validation of home-based sensor platforms (86, 170), despite the increasing commercial prevalence of such systems (171). The final study in this thesis aims to provide this validation; implementing a monitoring system based on techniques and principles studied in previous chapters.

7.1.2 A prototype system design

Through collaboration with a team of software engineers, offline functions developed to classify activity and measure movement parameters were integrated into a user-friendly end-to-end system. The system was designed to quantify execution of rehabilitation exercise,

measure movement parameters of specific exercises and provide feedback to both patient and clinician/investigator.

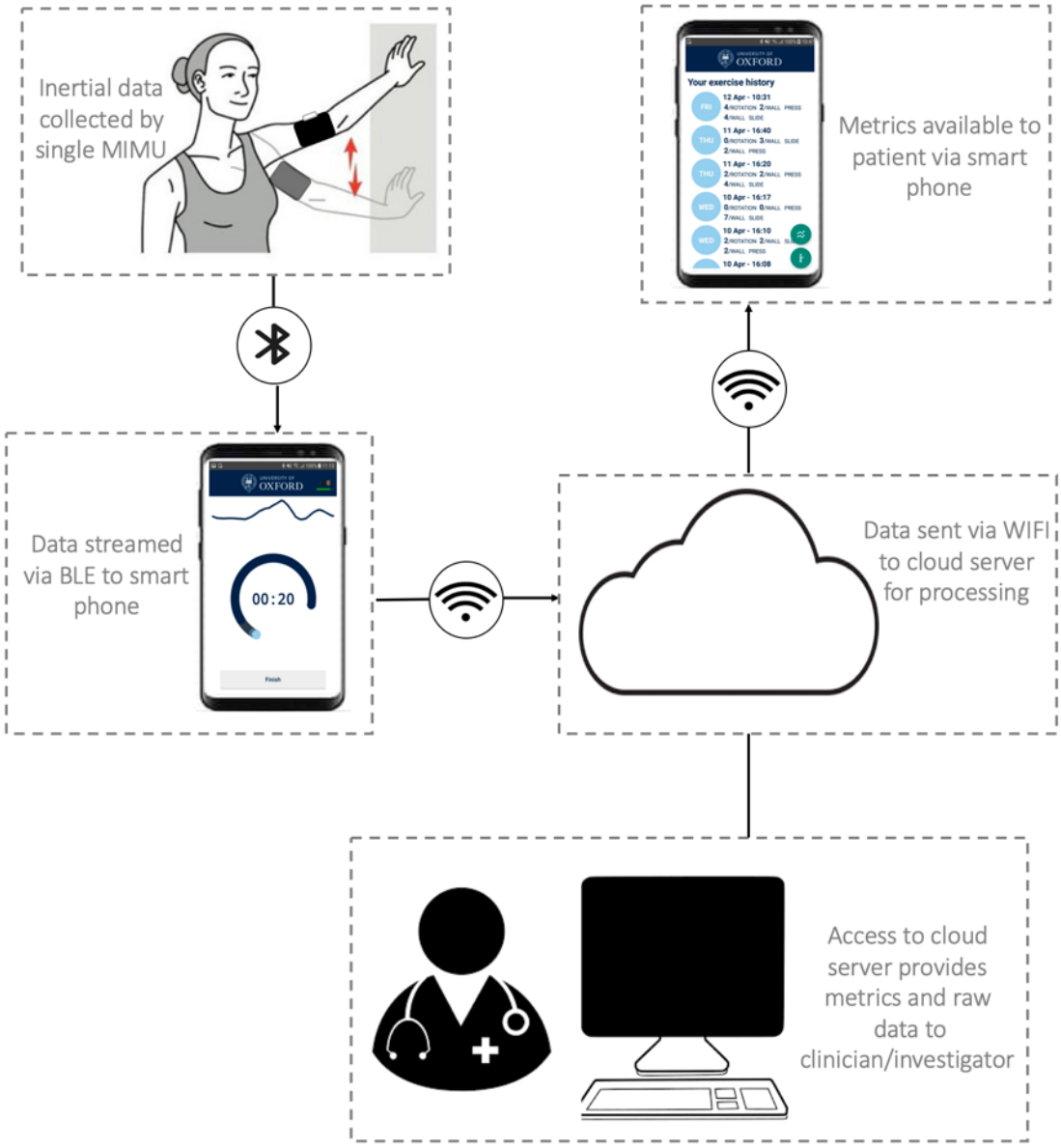


Figure 7.1. System architecture of the wearable motion capture platform.

Motion detection was achieved with a single inertial sensor (MetaMotion R, MbientLab, San Francisco, California, USA) mounted to an arm sleeve, and users were able to interact with the device and review recorded data via a custom-designed android-based

application. All three MIMU sensor modalities (accelerometer, gyroscope and magnetometer) were utilised and maximum sampling frequencies were chosen (100Hz, 100Hz and 25Hz respectively). The MIMU streamed raw inertial data to a smartphone via BLE, which was subsequently sent via a wireless internet connection to a secure cloud server. The raw data underwent a processing pipeline modelled on previous work. Output metrics were sent back to the user via the smartphone application.

No personal details were recorded on the device and users were registered with a four-digit identifier. The interface was designed to minimize user burden and to be suitable for individuals with any level of technological literacy. A single button was used to initiate or end an exercise session. Pairing with the sensor, sending and receiving data from the cloud, compilation and presentation of output metrics were fully automated. A graphic on the screen, shaped by live gyroscopic vector readings, indicated when data was being actively recorded and reflected the magnitude of the movement. Summary output data from each session was available to users via the same application and to investigators via an online server.

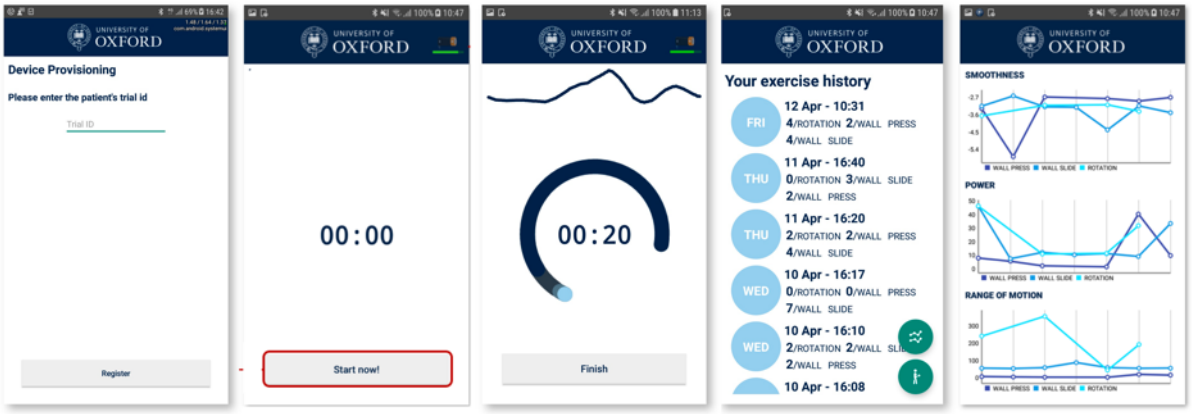


Figure 7.2. User interface (a) when registering user, (b) initiating exercises, (c) during exercise recording, (d) for reviewing adherence data and (e) for reviewing performance metrics.

7.1.3 System output

Recognition and classification of exercise activity was the central function of the platform. This function allowed measurement of engagement with rehabilitation and facilitated measurement of performance specific to certain exercise activity. The user was able to track their progress by viewing exercise counts from previous sessions. Implementing the segmentation technique developed in Chapter 6, pre-processed data was divided based on periods of zero velocity. A random forests algorithm was used to classify inertial data segments into a specific exercise type or junk activity. Following automated classification of all segments from the data stream, non-specific activity was disregarded and only relevant exercise segments were registered. With access to the cloud processing environment, it was possible to update the classifier at any time for a version trained using the latest user's data.

Performance metrics were calculated for classified data segments that represented specific exercises. This was an automated process and did not require any specific intervention from the user, such as having to indicate when specific movements are being performed or needing to strictly control recording of data. The parameters derived from the motion data included RoM, smoothness and power. RoM can provide a direct measurement of functional ability and is considered to be a good indicator of response to a rehabilitation program (63). Smoothness and power are novel metrics validated in previous chapters for quantitative evaluation of clinical progress in subacromial shoulder pain patients undergoing rehabilitation. Serial recordings of these movement parameters could provide a robust and efficient way to monitor clinical state through the rehabilitation process.

7.2. Study aims

The aim of this study was to investigate the properties of the prototype wearable motion capture system, including quantifying the accuracy and consistency of the output data, and understanding factors related to the usability and perceived impact of the system. The study was conducted with a view to understanding the feasibility of employing wearable systems for measuring the implementation and progress of home rehabilitation for clinical and research purposes.

7.3. Methods

7.3.1 Processing code development

The processing pipeline used to convert raw inertial sensor data to compliance metrics and performance scores was based on the MATLAB code written for the work carried out in previous chapters. Performance score calculations were performed on all data segments within the for-loop used to generate features for classification. Once the classifier determined which data segments represented each exercise activity, their respective performance scores were automatically retrieved and compiled within a numeric matrix. Scores calculated for non-exercise activity were discarded. Outliers were subsequently removed from the matrix and mean scores calculated for each exercise for the session. This summary score along with an exercise count was presented to the patient via the applications. The full processing pipeline is provided in Appendix D.

7.3.2 Study participants and setup

This study was granted ethical approval by a UK NHS Research Ethics Committee (17/SW/0217). Patients with a diagnosis of subacromial shoulder pain and who were deemed suitable for outpatient physiotherapy were recruited from NHS elective orthopaedic clinics. Criteria for exclusion included adhesive capsulitis, inflammatory arthropathy, any neuromuscular condition, cervical spondylosis or a history of recent trauma. 15 patients (eight males, seven females) meeting the eligibility criteria provided informed consent for their participation in this study. Four patients had partial or full degenerative rotator cuff tears, and six patients had previously received some form of treatment in the form of physiotherapy or steroid injections. Mean participant age was 54 years (SD 12.4), and 11 participants were right arm dominant.

A baseline assessment was performed after the study information was provided, eligibility confirmed and consent obtained. This assessment included collection of demographic data, and completion of the OSS. A user-training session of the sensor system was then provided, which was coupled with the instructions and guidance on how to carry out a recommended home exercise regime (Figure 7.3). Patients were provided with an information booklet detailing the exercises as well as a link to online information and videos. Participants attempted to carry out all exercises under supervision and while using the wearable sensor. This provided an opportunity to ensure patients were capable of performing the exercises correctly and to provide feedback on performance. Additionally, the inertial data recorded during the session provided training data for the classifier and a baseline assessment of function.

Patients were advised to follow the rehabilitation program guidance and to use the wearable device whenever their exercises were performed. Follow-up assessments were scheduled four to six weeks after baseline assessment, which included completion of the OSS and a feedback questionnaire. Patients also underwent a semi-structured interview and were asked to complete the System Usability Scale (SUS).



Figure 7.3. Subacromial pain rehabilitation guide published in 2018 by British Shoulder and Elbow Society with support from the NHS Get it Right First Time (GIRFT) Programme.

7.3.3 Evaluation of activity recognition

The classification pathway, including pre-processing, segmentation and feature selection, was modelled directly on the pathway that underwent cross-validation testing in Chapter 6. Data previously acquired from 20 patients and 20 healthy individuals, as well as data collected from each study participant during their baseline assessment was used for model training. During baseline assessment, a log was kept of exercise timing, sequence and count, which was used for labelling training data sets. After each subject was recruited, the classifier on the cloud server was replaced with an updated version trained with data from that subject.

Testing was performed only on data collected in the home environment during an unsupervised exercise session. The final home data session provided by each participant was used for testing to ensure the testing sets reflected the natural variations in exercise performance seen over time. Exercise type was manually verified through inspection of orientation data traces. Overall accuracy of classifier was reported along with individual exercise sensitivity, specificity and PPV. An accurate exercise count was dependent on accuracy of classification as well as consistency of segmentation. Therefore, percentage error in exercise count was also calculated.

7.3.4 Performance scores

Each exercise session yielded a summary score for each metric (RoM, smoothness and power), based on an average of all repetitions for each exercise after removal of any outliers. To explore the relevance of these scores and their potential to provide insight into patient recovery, trends over time were qualitatively reviewed alongside patient reported measures of pain and function. Statistical validation of these metrics was not possible due to the sample

size and lack of control group. This feasibility data was collected with view to planning a more comprehensive evaluation of serial measurement of performance in rehabilitation.

7.3.5 Usability and perceived impact

To provide quantitative data to support the evaluation of the wearable system, participants completed the SUS at follow-up review. The SUS is used as a self-reported measure of the perceived usability of the platform. It can be adapted to assess a range of systems, including mobile applications and websites, and consists of ten questions with answers provided on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree) (172). The output from each questionnaire is then used to generate a score out of a 100. Usability can be benchmarked against scores available from a growing body of literature evaluating systems' usability performance. It has been demonstrated that the SUS is a reliable and valid measure of perceived usability and can effectively distinguish between usable and unusable systems. The SUS has also been shown to be reliable at measuring usability with small sample sizes (173, 174).

A feedback questionnaire and a semi-structured interview was also used to provide qualitative information on usability and the perceived impact of the technology. The interview included open ended questions to explore the user experience of the system, and to gather information on how the system could be improved. To ensure there was consistency across each interview, questioning was structured according to a prepared topic guide (Appendix D). The topic guide was created based on the aims of the study and overall goals of this research.

7.4 Results

Use of the wearable system across subjects varied from four to 32 exercise sessions in total. Three participants did not complete more than five sessions, reporting excessive pain to have restricted their ability to exercise. A system bug resulted in processing of incomplete portions of certain session recordings. This was identified at an early stage of recruitment and corrected with an application update.

7.4.1 Accuracy of activity recognition

2,726 data segments from 15 exercise sessions (one per participant) were used for testing the accuracy of the classifier. Exercise activity labels were fairly evenly distributed; 607, 622 and 603 segments represented wall slide, wall press and rotation activity, respectively. Junk activity formed the most abundant class with 894 segments, representing 33% of all data segments. The overall accuracy of the classifier was 94.6%. Sensitivity, PPV and specificity was highest for wall slides and shoulder rotation exercises. Sensitivity for recognising wall press activity was lowest (81.0%), as was PPV for junk activity (87.0%). This reflected the tendency for wall presses to be misclassified for junk activity, which can be better appreciated in the confusion matrix (Figure 7.4).

	Overall Accuracy	Sensitivity	PPV	Specificity
	94.6 (±7.9)			
Wall Slide		99.3	99.3	99.8
Wall Press		81.0	97.3	94.6
Rotation		98.0	98.3	99.4
Junk		97.3	87.0	98.6

Table 7.1. Classifier sensitivity, PPV and specificity for each activity class.

Exercise count was generally underestimated for most subjects and the overall average discrepancy equated to a 10.3% underestimation in number of repetitions counted. Average discrepancy for each exercise per patient ranged from 7.7% overestimation to 84.6% underestimation. Both misclassification and under-segmentation was found to contribute to the underestimation in exercise count.

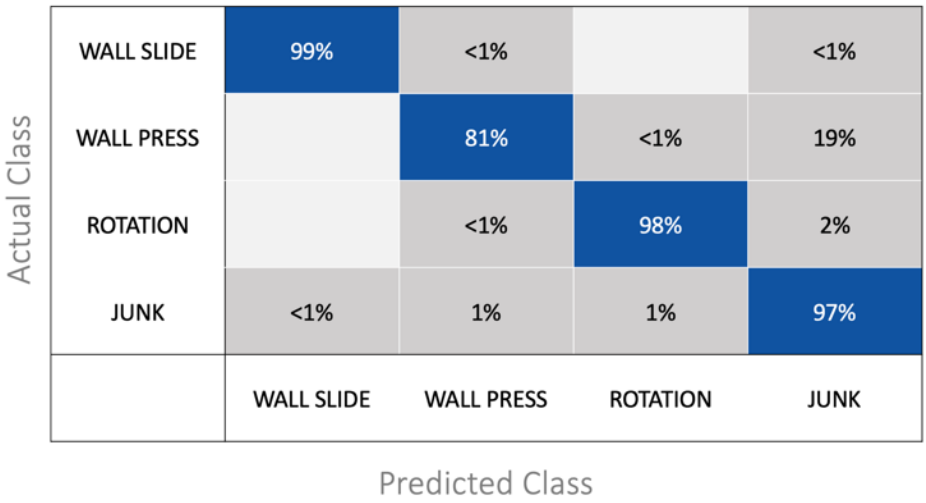


Figure 7.4. Confusion matrix for classifier testing results.

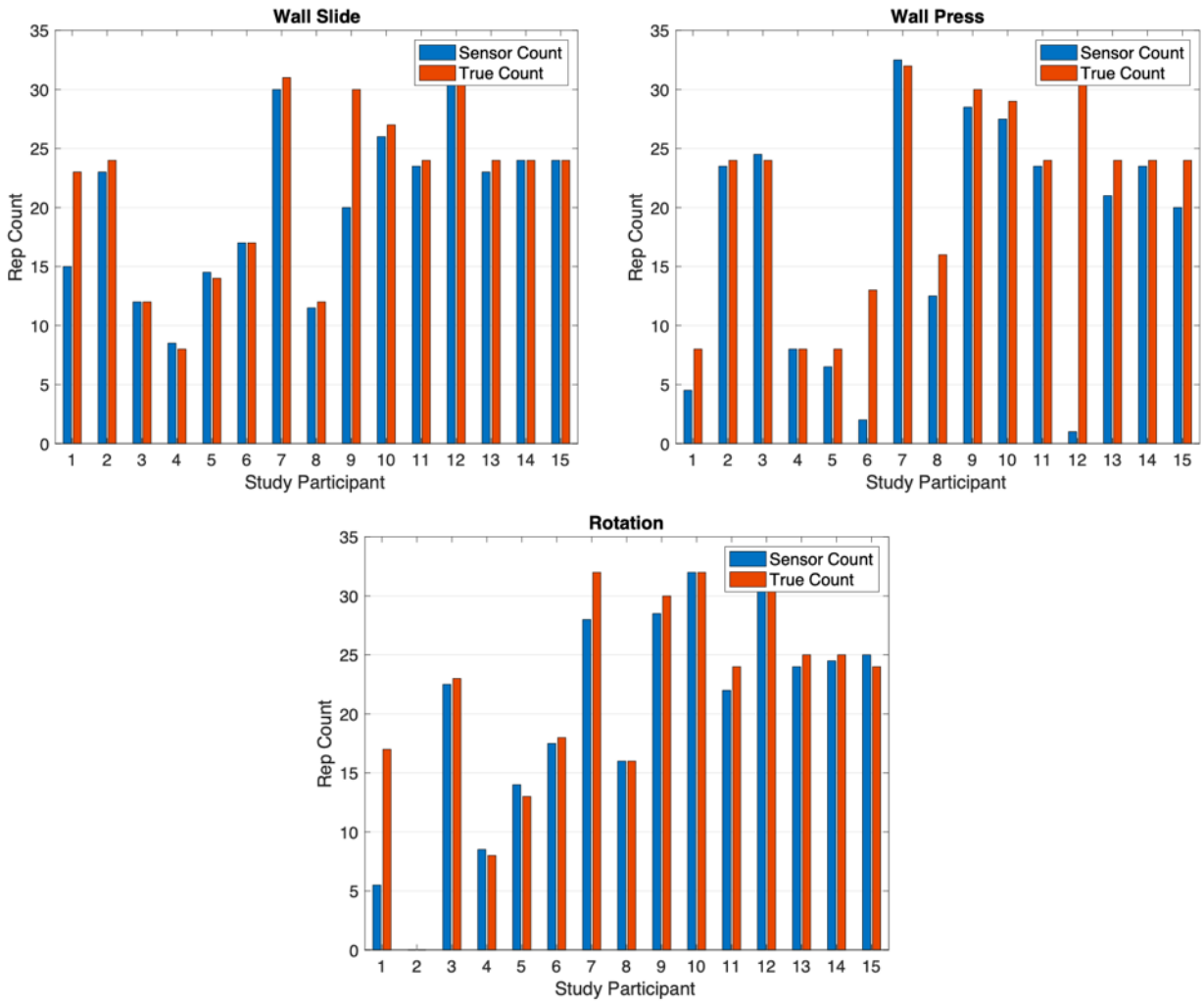
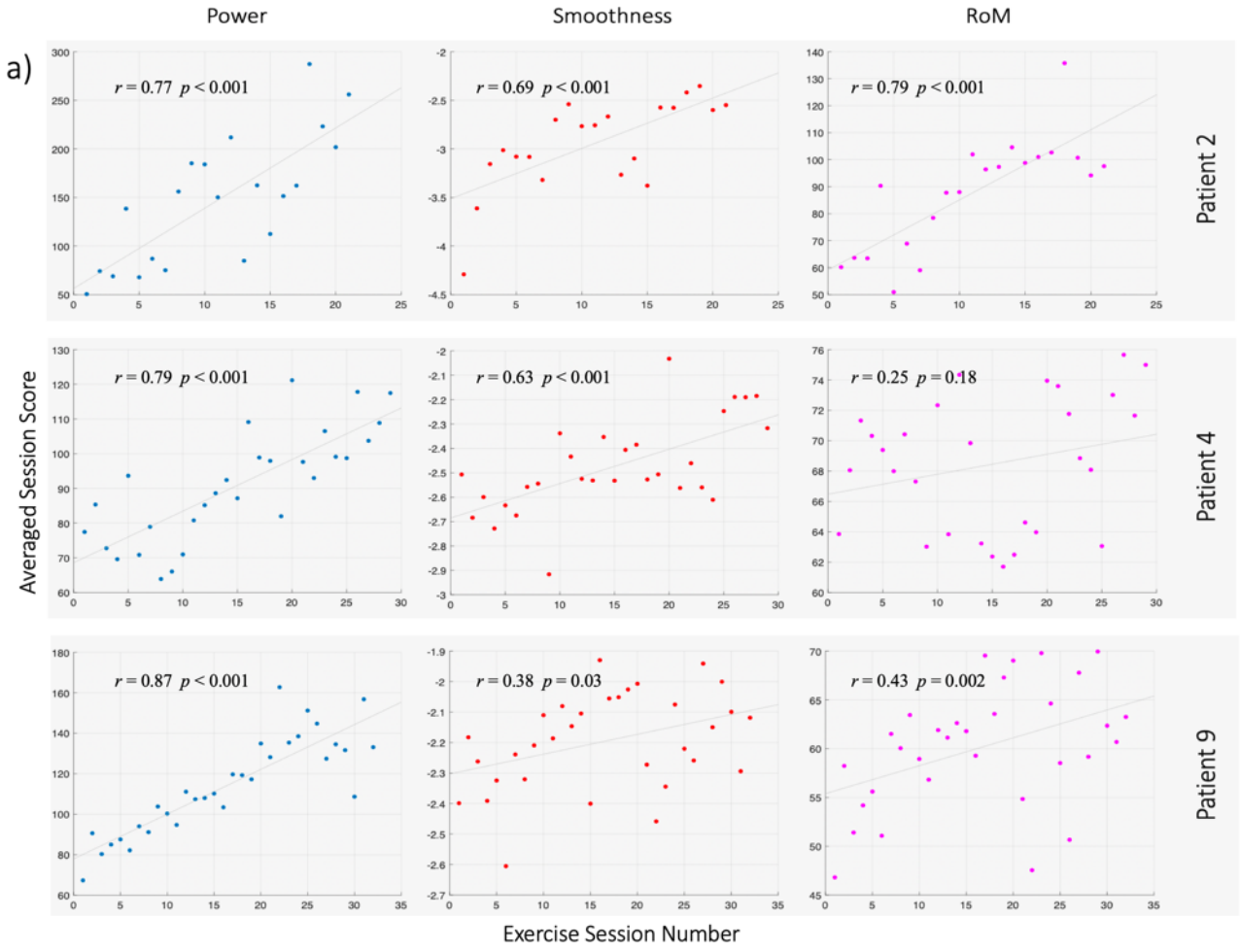


Figure 7.5. Bar charts for each exercise comparing exercise count from the wearable sensor to true number of repetitions performed.

7.4.2 Performance score profiles

Patients who engaged with their rehabilitation and reported clear improvements in their symptoms via the OSS, appeared also to show improvements in their score profiles over the study period. These patients demonstrated increasing trends in smoothness, power and RoM, quantified with Pearson's correlation coefficients, which were found to be statistically significant. Patients who reported difficulty engaging with their rehabilitation and struggled with pain often had erratic score profiles with no significant trend. The plots in figure 7.6

demonstrate select patients' scores for all recorded sessions (averaged across all 3 exercises) for smoothness, power and RoM.



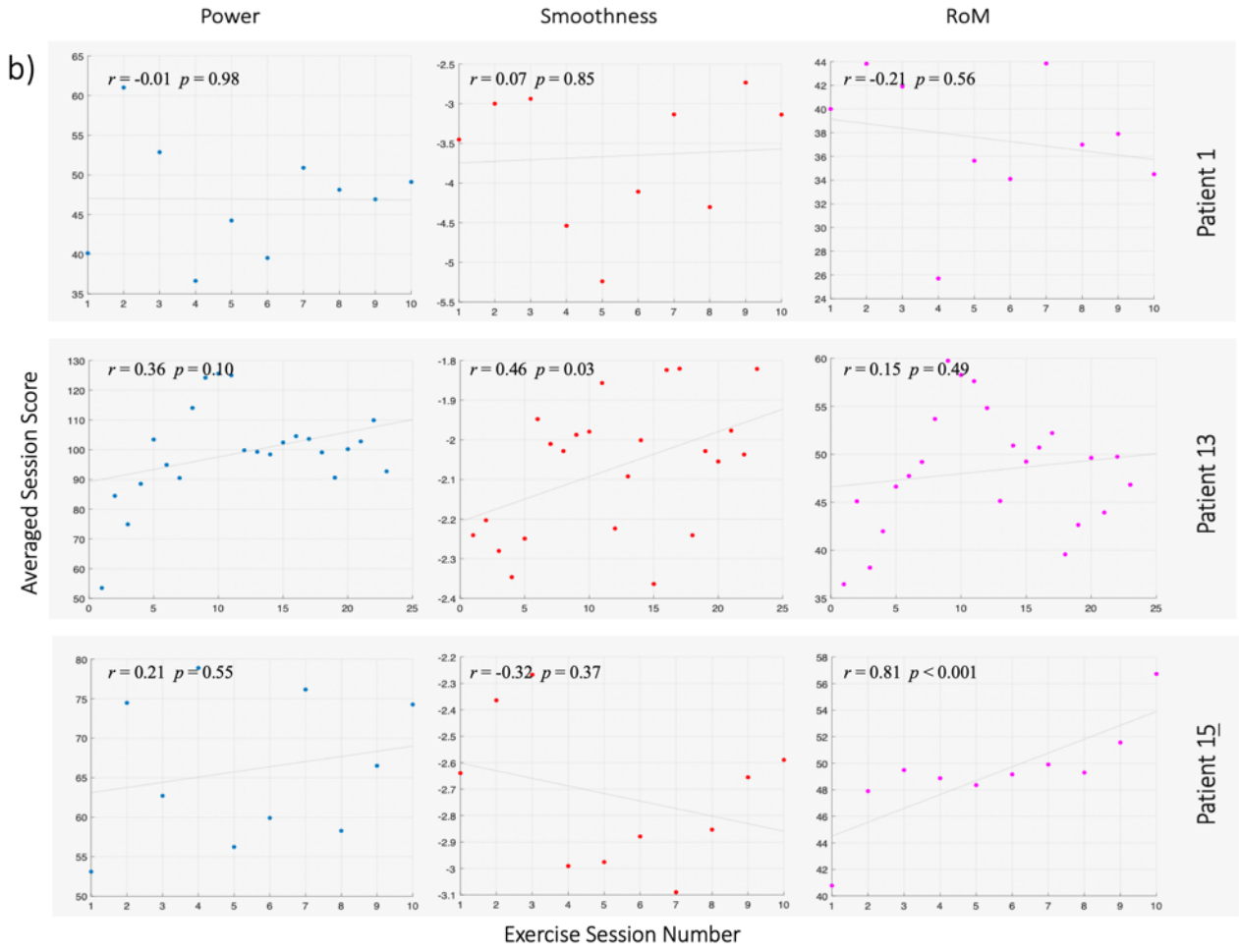


Figure 7.6. Power, smoothness and RoM scores following all rehabilitation sessions for select (a) patients with reported improvements and (b) patients with no reported improvements in pain and function.

7.4.3 Usability

The mean SUS score was 80.2 (SD 8.2) and individual scores ranged from 57.5 to 90. This score places the system in the 89th percentile when compared to other systems which have been evaluated using this scale (175). Although not designed to diagnose specific usability problems, it should be noted that scored responses were particularly low for question 6 ("I thought there was too much inconsistency in this system"), with an average of 2.1 points of the potential 4 points given for each item.

Qualitative data from the questionnaire and semi-structured interview provided greater detail on specific aspects of usability. All subjects, except one, agreed or strongly agreed with the statement that "the instructions for applying and setting up the sensors before exercises were easy to follow". Several subjects commented on the ease of system setup prior to exercising. Example statements included the following:

> *The armband was very easy to put on, but a few times I had to check the guide to make sure it was the right way up. The app was very simple. With only a few options to choose from, it was hard to go wrong.*

> *I liked how simple the app was, there's no need to overcomplicate things. Armband was easy enough.*

All patients disagreed or strongly disagreed when asked if the sensors caused discomfort or restriction on exercise performance. Patients were generally positive about the wearable and how it felt:

> *I liked the fact that I could stick it on over my clothes. It meant I could do the exercises anytime it was convenient.*

> *I wasn't really aware of it when I got going with my exercises. Definitely wouldn't say it was uncomfortable.*

However, two participants did report problems with performing shoulder rotations due to the positioning of the sensor:

I needed to rest my arm on a cushion while performing the rotations, and the sensor would catch whenever my hand was in the air, it was fine for the other exercises.

Several participants discussed the inconvenience of having to use a separate device for exercise monitoring, stating they would have preferred to install the application on their personal smart device:

I wish I could have downloaded the app on my phone. That way I wouldn't have had to worry about charging another phone and connecting it to WIFI wherever I went. If it was on my own phone I would be more likely to do the exercises, simply due to the convenience.

I'm not used to using android phones, so navigating around was a bit of a fiddle. If it was on my phone I would have been comfortable with it immediately.

All patients agreed they could see the benefits of the system for a younger population undergoing rehabilitation. However, three participants did not feel the older population would benefit and a two provided neutral responses to this question. While this issue is related to impact of the technology, potential issues with usability were quoted as underlying reasons for this:

Despite the simplicity, I don't see my elderly family members caring enough to go through the hassle. It still involves using a smart phone and connecting to the internet.

If elderly people are less likely to use emails and online banking, I suspect they'll be less likely to use this.

14 participants reported that they were willing to use the system again for rehabilitation. One subject offered a neutral response to this question, but did not report any specific issues or concerns with usability.

7.4.4 Perceived impact

All 12 participants who completed more than five home exercise sessions felt that the system improved their engagement with the rehabilitation program. There were different underlying reasons for this:

It was quite nice to see how consistent I became with my exercises and how the number of wall presses increased as I got better, there was a real sense of reward. The fact that it was being measured made me take it more seriously.

To be completely honest, I felt I needed to exercise because you would see exactly what I was doing. It's a simple way to motivate someone, and it works.

Several patients commented on how attempting to log exercises through the classifier encouraged a more complete performance of the exercises, focusing their efforts on quality rather than quantity.

I noticed if I was in pain or got tired, the sensors just wouldn't pick them up, clearly because the movements weren't big enough or they weren't being done properly. This was a bit frustrating at first, but I just made sure I did a few exercises properly and tried to improve each day.

It didn't quite capture them at first but I realised I wasn't doing the movement correctly. It was reassuring to know when I was doing them properly.

Another perceived benefit included the ability to share potentially clinically relevant motion data with a health professional. 11 subjects agreed or strongly agreed with the statement that "this system could help my doctor/physiotherapist understand my problem better". Only four patients found the performance metrics useful, most of the remaining patients did not feel the metrics enhanced their experience in any way.

It was great to see the graphs steadily going up. It doesn't give me something to aim for when I'm doing the exercises, but it's helpful to know I am better now than I was before. And that's certainly how I feel.

I did notice the numbers really dip when I struggled, but I can't say I found it helpful to watch them. They probably mean more to you than they do to me.

Most patients reported some discrepancy between the number of repetitions they performed and what was counted by the classifier, and this was found when evaluating the classifier. These patients understood that recognition was dependent on quality of exercise performance, but reported that feedback as to why exercises weren't being detected would have

improved their experience. This is discussed in greater detail in the following section. Patients did not express any significant concern regarding security of their data or breach of their privacy. All but one patient disagreed or strongly disagreed with the statement that they "would be concerned with how their motion data is used".

7.4.5 Improvements

Participants offered several suggestions for improvements and the most popular category related to quality and delivery of feedback. Some patients wanted a more real-time assessment of movement, and specific feedback aimed at improving performance.

> *I wish it would tell me why some exercises weren't being picked up. Or it would be helpful if the system was telling me what exercises are being logged as I have done them. With that kind of direct feedback, I would at least know what exercises I am doing correctly.*

> *It wasn't picking up my rotations until I realised it was because my elbow was too low down. I wish the system had told me that and maybe brought up the guidance information from the website.*

> *It would be helpful if there was something on the screen to show if my arm was in the right position or if I could improve what I was doing.*

Other suggestions for improvement included having an indicator on the arm sleeve to show correct sensor orientation, relocating the sensor to the lateral side of the upper arm and providing the ability for users to report their symptoms and experience through the application.

7.5 Discussion

7.5.1 Implications of the study

This study was a mixed methods evaluation of the use of a wearable motion capture system for home rehabilitation monitoring. Employing efficient kinematic assessment with a single MIMU sensor and online cloud processing, the system was designed to quantify several domains of exercise adherence and performance with minimal user burden. Quantitative assessment of usability demonstrated that patients were largely comfortable with this form of monitoring. The SUS score placed the system in the 89th percentile, which is comparable to other biofeedback systems for rehabilitation (171, 176, 177). Qualitative data collected from questionnaires and patient interviews also suggested a high level of usability for this prototype system. In addition to this, many patients highlighted the positive impact the system had on their compliance and engagement with treatment. Patients also saw value in their motion data being shared with a clinician, potentially enabling a better understanding of their condition and how they are responding to treatment. It was clear that usability and perceived value could be enhanced with improvements in system accuracy and technical developments in feedback delivery. Further developments will need to maintain an appropriate balance between usability and functional capability.

Identifying the active ingredient of any rehabilitation program is a challenging task and despite the dedicated research into determining the type, quantity and intensity of exercise required to achieve optimal outcomes, there is little evidence to support one regime over another (9, 10). Inertial sensors can address this problem through accurate measurement of an extensive range of movement parameters during exercise treatment. Previous studies have demonstrated the technical feasibility of using machine learning techniques to classify rehabilitation exercises. However, evaluation in natural settings, alongside usability assessment of an end-to-end platform is necessary to determine feasibility of implementation in real world scenarios. This study demonstrates that accurate compliance metrics can be generated from a single sensor in a practical and unobtrusive way. Quantifying treatment fidelity in this level of detail will allow for a more careful and informed interpretation of rehabilitation treatment outcomes.

Assessment of recovery is an important aspect of any rehabilitation program. Most methods to achieve this do not provide objective data on physical movement characteristics. This data may add value in monitoring patient progress and properties of a rehabilitation program. Within this feasibility study, patients with positive outcomes had inertial scores that appeared to show greater consistency and a tendency to increase over time. The findings suggest there is strong potential for this form of serial assessment for routine clinical monitoring and for evaluating response to treatment. Patients who are struggling with their rehabilitation and failing to progress can be identified promptly and efficiently. The role of connected health solutions for safety netting patients in the community is well established (17), and these systems can improve the efficiency of clinical service delivery, identifying those most in need. Sufficiently powered, longer term, controlled studies will be required to further investigate and validate this form of assessment.

7.5.2 Summary

MIMU based biofeedback systems not only offer a reliable approach to measuring exercise performance and adherence, but can also address factors that influence adherence through self-monitoring, providing automated feedback and goal-setting. The findings of this study will inform the design of future systems, with the aim of improving usability and encouraging user-engagement. Potential developments to the existing prototype system and future directions for research implementing MIMU based wearable platforms are discussed in the final chapter.

8 Discussion

Contents

8.1 Summary of main findings and conclusions 148
8.2 Limitations 151
8.3 Recommendations for future work 153

8.1 Summary of main findings and conclusions

The goal of subacromial shoulder pain management is to improve the symptoms of pain and restore function of the shoulder. The lack of consensus on precisely what form of rehabilitation can achieve this (10, 14), raises concerns over the clinical and cost effectiveness of most exercise-based treatments. To understand the relationship between rehabilitation strategies and their potential outcomes, it seems essential to have detail on dosing and timings of the interventions. As with pharmaceutical agents, an understanding of the active ingredient, target and effect should be known for rehabilitation treatments. A failure to measure these properties and quantify treatment fidelity has undermined the development and testing of effective exercise programs, and prevented the potential success of treatments implemented in clinical trials (178). Limited means to accurately and objectively quantify physical impairment and its impact on clinical function (179), also hinders our ability to measure the efficacy of exercise-based treatments. Developments in connected health solutions, utilising the accessibility of mobile devices and practicality of micro-electromechanical systems (MEMS) based sensing technology, can help overcome these obstacles to deliver purposeful and effective rehabilitation. Therefore, this thesis aimed to:

Develop a MIMU based motion capture system for the shoulder and investigate its practicality in providing meaningful and useful feedback on shoulder movement and shoulder rehabilitation.

The objective of the first study (Chapter 3) was to develop a pre-processing pathway, capable of deriving reliable measurements of acceleration, speed and orientation from raw MIMU sensor data. Testing of a specific MIMU sensor using a custom-built pre-processing

pipeline was performed using a robotic platform that generated a range of measurable dynamic movement conditions, which were controlled and repeatable. The following study in Chapter 4 evaluated the accuracy and reliability of a sensor mounted onto an ergonomically designed wearable to reference 3D orientation of the upper arm segment. This involved testing with human subjects, against a gold-standard optical motion capture reference system evaluating shoulder movement in three DoF. Chapter 5 explored the potential for alternative MIMU sensor derived kinematic metrics to assess clinical function. The study investigated the psychometric properties of these metrics in the evaluation of patients with subacromial shoulder pain. Developments to the original processing pathway were introduced in Chapter 6, to facilitate segmentation and classification of activity data streams. This approach to measuring exercise adherence was validated using data from a cohort of subacromial pain patients performing a set of typical shoulder exercises. The MATLAB functions within the final processing pathway were then incorporated into an online back-end cloud computing platform, which communicated with the user and sensor via an android based custom designed application. The final study (Chapter 7) trialled this prototype end-to-end system with patients undergoing home rehabilitation for subacromial pain, to evaluate system accuracy, usability and perceived impact of the technology.

Through development of innovative data processing techniques, novel application of existing programming functions and testing under clinically relevant conditions, this thesis has made several important contributions to the field of study:

An orientation filter based on a gradient descent algorithm was found to show levels of dynamic and static accuracy suitable for clinical motion analysis. The performance demonstrated was comparable to reports of more established filtering methods. Accuracy will

naturally be influenced by several other variables, including choice of MIMU, pre-processing methods and testing conditions. However, this investigation was the first to demonstrate a reliable influence of movement speed and direction on this filter's accuracy, and a stable level of performance over time. The study provides translatable findings for researchers and clinicians aiming to employ computationally efficient orientation filtering for near real-time clinical applications.

Metrics derived from an upper arm MIMU-based kinematic data show construct validity in the assessment of patients with subacromial shoulder pain. Movement smoothness calculated from inertial data was formally evaluated for the first time in patients with musculoskeletal pathology, and scores demonstrated reliability, known group validity and responsiveness to treatment. Metrics based on ranges of accelerometer and gyroscope data also showed similar psychometric properties, as well as convergent validity, demonstrated by significant correlation with a validated PROM scores. Significant effect sizes for score changes of these metrics following treatment were also seen over the 12-week study period. This form of assessment proved to be quick and practical, and shows clear potential to complement existing methods of clinical assessment.

A fully automated coding pathway was created and validated for recognition and classification of shoulder exercise activity using a single inertial sensor and supervised machine learning techniques. Chapter 6 introduced novel processes to facilitate dynamic data segmentation and to exclude non-specific activity; both important requirements to providing a meaningful assessment of exercise activity. This work was the first to validate these processes in patients with shoulder pathology undergoing treatment, and evaluate the use of an end-to-end clinical rehabilitation monitoring system in the home environment. Testing under these

conditions demonstrated that a practical wearable system is currently capable of providing a reliable and detailed account of treatment fidelity in real world settings.

Quantitative and qualitative data gathered in the final study contributes to the small but growing body of work exploring the usability and patient perceived value of sensor-based systems. This is a critically important area of research within the field of clinical wearables, as the potential benefits of biofeedback systems can only be realised if there is a sufficient acceptance and uptake of the technology. The study demonstrated that there was general acceptance of the prototype system and some participants reported that it had a positive impact on their treatment compliance. The findings of this study will inform the design of future systems, further improving usability and the potential for increasing engagement.

8.2 Limitations

A standard limitation, shared by inertial sensors and all other forms of motion assessment (based on surface mounted sensors or markers), is the noise created by movement of the soft tissues relative to the underlying skeleton. Attempts have been made to model the artefact for motion sensing applications, but no method can perfectly account for variations in soft tissue properties, fit and positioning of the wearable. Recommendations to reduce motion artefact include firm attachment with compression of the soft tissue, and strict standardisation of placement (180). Unfortunately, this would result in a wearable that is difficult to setup and uncomfortable to wear, particularly during the performance of demanding exercises. Regardless of how sophisticated processing methods become, comfort and *wearability* are likely to come at the cost of accuracy and reliability. Advancing the use and impact of wearable

systems will require challenging compromises to be made between usability and technical performance.

The clinical significance of the inertial metrics was evaluated against the Oxford Shoulder Score (OSS). The OSS is a validated patient reported outcome measure (PROM) for the assessment of patients with shoulder pain. PROMs are increasingly being used to improve patient care by measuring aspects of health that patients consider important. They provide insight into the efficacy and acceptability of medical and therapy-based treatments, and are increasingly utilized in routine clinical practice. However, most PROMs were developed and validated for use at a group level in research studies, with little validation for their use on an individual patient level (181). Patient-based assessment tools also suffer from bias, with patients under or over-reporting symptoms depending on their relationship with their healthcare provider and other factors unrelated to their treatment (182). PROMs are clearly valuable in capturing the patient perspective, but can correlate poorly with objective measures that quantify performance and are significantly influenced by pain (179). Assessment of patients' functional capabilities is a complex task, and the relationship between inertial scores and PROMs should be interpreted with caution.

The classification algorithms used for activity recognition operate by applying a set of rules to recognise patterns from the sensor data. These rules are independently established during model training with real patient data, rather than being manually programmed. However, the association between the activities and their MIMU data profiles is influenced by how the training data is collected and labelled, which introduces a level of subjectivity to the process. During labelling of the training data in Chapter 6, exercises that were not executed correctly were not included. Determining what constitutes correct performance is not a precise

science and is dependent on the abilities of the patient. A conservative approach to labelling would result in a classifier with unrealistically high expectations of the user, only logging exercises that have been executed perfectly. This could be disheartening for the patient and deny the clinician of a fair account of the patient's engagement. A more inclusive labelling approach could compromise overall accuracy by distorting the boundaries between classes, particularly between exercise activity and non-exercise activity. This approach could also inappropriately reward patients for poor exercise performance. A perfect system would classify all exercise attempts as their intended activity and objectively quantify the performance of execution. This would certainly be most useful for both the clinician and patient, but it is unlikely that this can be achieved with a single sensor platform. Once again, there is a trade-off between ergonomics and function, and the setting of technical goals going forward will be a difficult task.

8.3 Recommendations for future work

Instrumented assessment of upper limb activity can provide valuable insights into clinical function and facilitate tracking of patient recovery with minimal investment of healthcare resources and time. Inertial scores calculated from two-minute sensor recordings underwent validation testing with promising results, and the final study demonstrated the feasibility of implementing longitudinal kinematic assessment for functional monitoring. Further validation involving larger sample sizes and longer follow-up periods will be required to understand the clinical relevance of this kinematic data. The scores performed well in differentiating pathological from healthy shoulders, but the ability to stratify patients based on functional ability or condition severity should also be explored. Significant score changes were seen following treatment, but we do not know what changes constitute treatment success. It

will be necessary to establish a minimal clinically important difference (MCID) to represent the smallest score change that signifies a clinically relevant change in symptoms. Other score characteristics which will need defining include possible floor or ceiling effects, temporal trends and the relative influence of pain and baseline characteristics.

The activity recognition pathway developed and tested in this thesis can provide researchers and clinicians with a convenient tool to objectively quantify rehabilitation treatment fidelity. However, further improvements in classification accuracy should be sought, to offer a more effective clinical tool, and also enhance user satisfaction and patient engagement. As discussed in the limitations section, there is a degree of subjectivity involved in labelling activity during classifier training. Further work in this area should be focused on establishing more objective approaches to data labelling and should also aim to quantify the consistency of these methods. A greater understanding of sensor noise and error will inevitably improve the quality of features derived from inertial sensor data. Improved methods of feature selection, such as the computationally demanding wrapper methods discussed in Chapter 6, should be evaluated (155). *Feature transformation* methods, using techniques such as principle component analysis, should also be considered for solving these activity classification problems. The new features generated using these transformation methods may not have the same interpretation as their original counterparts, but can have greater discriminatory power (183). In addition to feature selection, model *hyperparameter* optimisation was not exhaustively investigated. In the case of the random forest algorithm, this would involve determining the optimal number of decision trees in the forest and the number of features considered when splitting a node. Optimising these parameters can have a significant impact on model performance. Finally, the use of deep learning techniques should be explored; neural network-based algorithms offer significant performance advantages over traditional supervised

machine learning techniques, and are proven to be very effective for exercise classification in healthy subjects (88).

There are evidently many options for improving the technical capabilities of MIMU based motion capture systems, some of which have been discussed. As well as providing an objective tool to assess and monitor patients, wearables have also demonstrated the potential to positively influence behavior and improve patient engagement with treatment (176). There is a clear role for wearable systems as an intervention; one that aims to improve quality and quantity of self-directed treatment by feeding back to patients and by providing a sense of accountability. In the context of general physical activity monitoring, providing visual feedback to users has been associated with greater exercise compliance, when compared to those who are monitored without receiving any feedback (184). This thesis demonstrates the viability of implementing wearable systems for home rehabilitation, and future work should evaluate the impact this intervention may have on engagement and most importantly, patient outcomes. Comparing the clinical and cost-effectiveness of sensor-supported rehabilitation to conventional treatments should ultimately be performed in the context of appropriately powered randomised controlled trials. Given the nature of the intervention, there will be challenges with many aspects of study design. For example, blinding will not be possible, and measuring treatment fidelity in the control arm will be problematic. Therefore, further feasibility work will first be required to gain a better understanding of the practical and operational issues of implementing wearables in clinical trials.

Musculoskeletal conditions are characterised by limitations in movement and restrictions in functional ability. Wearable technologies have the potential to revolutionise musculoskeletal care by offering improved ways of collecting, processing and presenting

human motion data. The possible applications of connected health solutions are broad and can have utility in many areas of clinical assessment and management. Motion data collected in natural environments can also offer new insights into the natural history of disease and individual responses to treatment. A great deal more research and development is required to harness the true potential of these biofeedback systems, and address significant challenges around measurement validity and usability. Moving forward, more multidisciplinary collaborations are needed to approach these challenges in a way that ensures clinical wearables can become a valuable asset for healthcare in the future.

Appendix A

Contents

A.1 Device specifications 158
A.2 Orientation error data 159

A.1 Device specifications

Dimensions	Length 29mm, width 18mm, height 6mm
Weight	2.8g
Connectivity	Bluetooth LE 4.2 – 2.4Ghz Typical range 10m Stream sensor data up to 100Hz
Battery	100mAH micro-USB rechargeable lipo battery
Accelerometer	±2g / ±4g / ±8g / ±16g Noise density: 180μg/√Hz
Gyroscope	±125°/s /±250°/s /±500°/s /±1000°/s / ±2000°/s Noise density: 0.008°/s/√Hz
Magnetometer	±1300μT (x,y-axis) range ±2500μT (z-axis) range Resolution: 0.3μT

Table A1. Metamotion R technical specifications. MetaMotion R, MbientLab, San Francisco, California, USA.

A.2 Orientation error data

Attitude Error											
		10 deg		20 deg		30 deg		40 deg		50 deg	
		SRE	SSE	SRE	SSE	SRE	SSE	SRE	SSE	SRE	SSE
0.2 Hz	Median	0.17	0.16	0.28	0.43	0.31	0.53	0.55	0.91	1.15	0.93
	95% UL	0.46	0.56	0.90	0.94	1.49	1.98	2.29	2.29	3.96	2.52
0.6 Hz	Median	0.22	1.40	0.37	1.68	0.55	2.46	0.82	2.12	1.19	1.40
	95% UL	0.62	2.86	0.97	3.34	1.39	7.09	1.92	5.39	2.62	5.00
1.0 Hz	Median	0.25	0.79	0.48	1.15	0.68	3.75	1.09	4.41	1.30	3.69
	95% UL	0.75	2.10	1.39	6.31	2.31	10.65	3.46	25.24	3.82	10.32
1.4 Hz	Median	0.48	0.80	0.93	2.93	1.57	3.52	1.91	3.92	2.76	3.86
	95% UL	1.37	2.41	3.13	11.04	4.81	9.71	4.87	10.20	7.40	16.24
1.8 Hz	Median	0.65	0.94	0.96	3.19	2.40	3.66	3.75	4.02	4.57	4.21
	95% UL	1.76	2.94	4.08	8.25	5.63	10.71	8.55	12.63	11.15	14.37

Heading Error											
		10 deg		20 deg		30 deg		40 deg		50 deg	
		SRE	SSE	SRE	SSE	SRE	SSE	SRE	SSE	SRE	SSE
0.2 Hz	Median	0.48	0.80	0.86	1.42	0.93	1.53	1.30	1.87	1.99	2.05
	95% UL	1.20	1.59	1.68	2.19	1.80	2.45	3.07	5.97	2.88	5.15
0.6 Hz	Median	0.43	1.07	0.68	1.10	0.91	2.42	0.99	2.49	1.15	2.70
	95% UL	1.12	2.64	1.89	3.15	2.37	6.72	3.20	6.67	2.75	7.27
1.0 Hz	Median	0.44	0.90	0.65	1.59	0.96	2.17	1.38	3.89	1.73	4.37
	95% UL	1.25	2.56	1.91	5.27	3.06	6.63	3.85	11.28	5.72	11.63
1.4 Hz	Median	0.43	0.88	0.66	1.21	1.22	2.66	1.74	4.83	2.54	6.50
	95% UL	1.32	2.40	2.00	3.93	3.69	7.65	4.98	14.11	6.93	15.89
1.8 Hz	Median	0.48	1.03	0.77	2.00	1.33	3.58	1.99	5.11	7.54	14.62
	95% UL	1.42	3.15	2.44	8.37	4.16	10.51	5.90	16.50	23.23	41.53

Table A2. Attitude and heading axis median error and 95% upper limit. Sensor-to-reference (SRE) and sensor-to-sensor error (SSE) reported over 300min trials.

Multi-axis Movement Error											
		10 deg		20 deg		30 deg		40 deg		50 deg	
		SRE	SSE	SRE	SSE	SRE	SRE	SSE	SRE	SSE	SRE
Attitude											
1.0 Hz	Median	0.28	0.65	0.49	1.25	0.70	2.34	1.02	5.22	1.61	4.40
	95% UL	0.89	9.71	1.51	3.43	2.85	29.21	3.01	18.31	4.30	12.56
Heading											
1.0Hz	Median	0.60	1.10	0.75	4.17	1.18	1.30	1.94	9.49	2.34	9.14
	95% UL	1.69	3.15	2.54	8.08	8.15	9.68	5.73	23.88	7.52	19.19

Table A3. Multi-axis median error and 95% upper limit. Sensor-to-reference and sensor-to-sensor error reported over 300min trials.

T1 vs T2 Error											
		10 deg		20 deg		30 deg		40 deg		50 deg	
		T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
Attitude											
SRE 1.0Hz	Median	0.25	0.25	0.47	0.48	0.69	0.67	1.11	1.07	1.23	1.38
	95% UL	0.74	0.75	1.51	1.29	3.07	2.04	3.27	3.74	3.62	4.03
Attitude											
SSE 1.0Hz	Median	0.82	0.75	1.16	1.13	3.84	3.66	4.30	4.53	4.02	3.37
	95% UL	2.40	1.93	6.64	6.07	10.73	10.73	25.76	24.90	10.33	10.29
Heading											
SRE 1.0Hz	Median	0.45	0.43	0.65	0.66	0.95	0.96	1.42	1.34	1.72	1.74
	95% UL	1.26	1.24	1.96	1.86	3.21	2.92	4.00	3.66	5.05	6.48
Heading											
SSE 1.0Hz	Median	0.89	0.90	1.60	1.58	2.42	1.92	3.97	3.82	4.28	4.45
	95% UL	2.55	2.57	5.24	5.30	7.24	5.94	11.33	11.19	11.96	11.26

Table A4. *Early* (T_1) and *Late* (T_2) portions of 300 min 1Hz trials for attitude and heading axes. Median error and 95% upper limit reported for sensor-to-reference readings.

Static Error			
		SRE	SSE
Attitude	Median	0.03	0.05
	95% UL	0.16	0.19
Heading	Median	0.21	0.24
	95% UL	0.62	0.69

Table A5. Median error and 95% upper limit in attitude and heading planes during 60 min static trial.

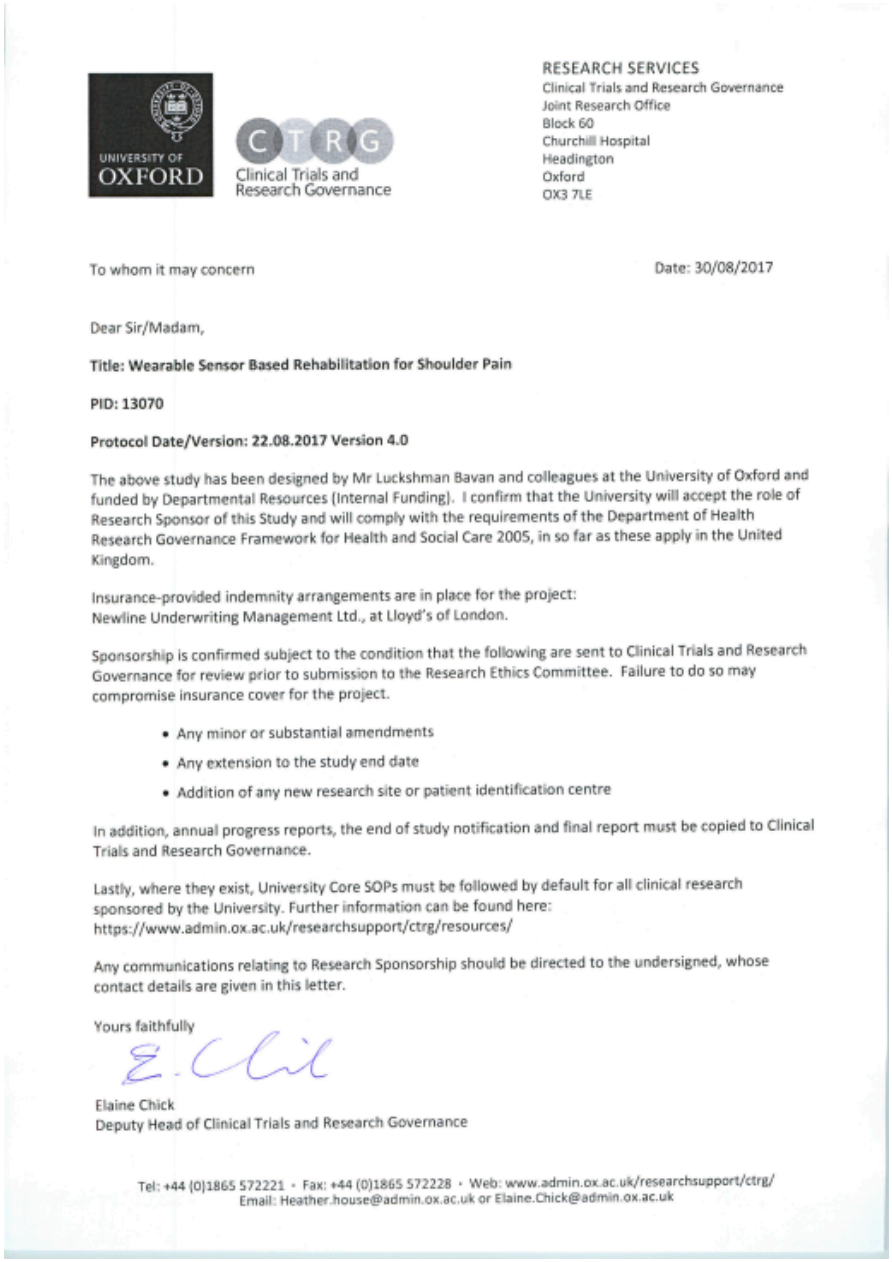
Appendix B

Contents

B.1 Ethical approval 163
B.2 Design of wearable accessory 168

B.1 Ethical approval

Ethical approval was granted for all studies in this thesis involving human participants. The University of Oxford (Clinical Trials and Research Governance) acted as the research sponsor for this work and Health Research Authority approval was issued following review by the Southwest-Central Bristol Research Ethics Committee.



UNIVERSITY OF OXFORD

CTRG Clinical Trials and Research Governance

RESEARCH SERVICES
Clinical Trials and Research Governance
Joint Research Office
Block 60
Churchill Hospital
Headington
Oxford
OX3 7LE

To whom it may concern

Date: 30/08/2017

Dear Sir/Madam,

Title: Wearable Sensor Based Rehabilitation for Shoulder Pain

PID: 13070

Protocol Date/Version: 22.08.2017 Version 4.0

The above study has been designed by Mr Luckshman Bavan and colleagues at the University of Oxford and funded by Departmental Resources (Internal Funding). I confirm that the University will accept the role of Research Sponsor of this Study and will comply with the requirements of the Department of Health Research Governance Framework for Health and Social Care 2005, in so far as these apply in the United Kingdom.

Insurance-provided indemnity arrangements are in place for the project:
Newline Underwriting Management Ltd., at Lloyd's of London.

Sponsorship is confirmed subject to the condition that the following are sent to Clinical Trials and Research Governance for review prior to submission to the Research Ethics Committee. Failure to do so may compromise insurance cover for the project.

- Any minor or substantial amendments
- Any extension to the study end date
- Addition of any new research site or patient identification centre

In addition, annual progress reports, the end of study notification and final report must be copied to Clinical Trials and Research Governance.

Lastly, where they exist, University Core SOPs must be followed by default for all clinical research sponsored by the University. Further information can be found here:
https://www.admin.ox.ac.uk/researchsupport/ctrg/resources/

Any communications relating to Research Sponsorship should be directed to the undersigned, whose contact details are given in this letter.

Yours faithfully

E. Chick

Elaine Chick
Deputy Head of Clinical Trials and Research Governance

Tel: +44 (0)1865 572221 · Fax: +44 (0)1865 572228 · Web: www.admin.ox.ac.uk/researchsupport/ctrg/
Email: Heather.house@admin.ox.ac.uk or Elaine.Chick@admin.ox.ac.uk

Figure B.1. Sponsor letter from the Clinical Trials and Research Governance team.

NHS
Health Research Authority
South West - Central Bristol Research Ethics Committee
Whitefriars
Level 3, Block B
Lewin's Mead
Bristol BS1 2NT
Email: nrescommittee.southwest-bristol@nhs.net

Telephone: 0207 104 8028
Fax:

Please note: This is the favourable opinion of the REC only and does not allow you to start your study at NHS sites in England until you receive HRA Approval

19 September 2017

Mr Luckshman Bavan
Botnar Research Centre
Old Road
Oxford
OX3 7LD

Dear Mr Bavan

Study title:	**Wearable Sensor Based Rehabilitation for Shoulder Pain**
REC reference:	**17/SW/0217**
Protocol number:	**13070**
IRAS project ID:	**228905**

Thank you for your letter of 15 September 2017, responding to the Proportionate Review Sub-Committee's request for changes to the documentation for the above study.

The revised documentation has been reviewed and approved by the sub-committee.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact please contact hra.studyregistration@nhs.net outlining the reasons for your request.

Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA Approval (England)/ NHS permission for research is available in the Integrated Research Application System, www.hra.nhs.uk or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" above).

Approved documents

The documents reviewed and approved by the Committee are:

Document	*Version*	*Date*
Copies of advertisement materials for research participants [Recruitment poster]	2	28 August 2017
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Clinical Trials Confirmation]		20 July 2017
IRAS Application Form [IRAS_Form_30082017]		30 August 2017
IRAS Checklist XML [Checklist_15092017]		15 September 2017
Letter from funder [Funding]		08 March 2017
Letter from sponsor [Letter from sponsor]		30 August 2017
Other [Response covering letter]		15 September 2017
Participant consent form [Consent form]	4	28 August 2017
Participant information sheet (PIS) [Patient Information Sheet (Tracked)]	6	15 September 2017
Participant information sheet (PIS) [Patient Information Sheet]	6	15 September 2017
Participant information sheet (PIS) [Volunteer PIS tracked]	6	15 September 2017
Participant information sheet (PIS) [Volunteer PIS]	6	15 September 2017
Research protocol or project proposal [Protocol]	4	22 August 2017
Summary CV for Chief Investigator (CI) [CV]		
Validated questionnaire [EQ5D]		
Validated questionnaire [Oxford Shoulder Score]		

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

Feedback

You are invited to give your view of the service that you have received from the Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:
http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance

We are pleased to welcome researchers and R & D staff at our RES Committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

17/SW/0217	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project.

Yours sincerely

pp. **Mr Brian Pixton**
Chair

Email: nrescommittee.southwest-bristol@nhs.net

Enclosures: *"After ethical review – guidance for researchers" [SL-AR2]*

Copy to: *Ms Heather House, Research and Development Department Oxford University Hospitals NHS Foundation Trust*

A Research Ethics Committee established by the Health Research Authority

Figure B.2. Health Research Authority letter of approval.

B.2 Design of wearable accessory

A waist clip was provided with each MIMU sensor purchased from MbientLab. This was pinned to a thin semi-rigid sheet of thermoplastic moulded to contour the posterior distal aspect of the upper-arm segment. This sensor mounting was then incorporated into a commercially available pocket arm sleeve (Ronhill Stretch Arm Pocket). Two sizes of plastic sheets were cut and moulded to match two sizes of arm sleeve: “small/medium” and “medium/large”.

In order to standardise sensor position relative to arm segment, users were advised to ensure that the exposed sensor surface is facing posteriorly when the arm is in a neutral position. The bottom of the sleeve should also line up with the elbow crease. The wearable could be worn directly or fitted over clothing.



Figure B.3. MIMU sensor incorporated into a wearable accessory.

Appendix C

Contents

C.1 Feature rankings 170

C.1 Feature rankings

Time Domain Features	Frequency Domain Features
Mean	Maximum Frequency Component
Root Mean Square	Mean Frequency Component
Standard Deviation	Energy
Variance	Entropy
Range	Excessive kurtosis
Inter-quartile range	
Percentiles (3,10, 20, 97)	
Correlation (xy/xz/yz)	

Table C1. Time and frequency domain features derived from accelerometer, gyroscope, magnetometer and orientation data segments. Features were calculated from all axes and normalised vector magnitude values of all sensor modalities.

Rank	Feature	Relieff Weight	Rank	Feature	Relieff Weight	Rank	Feature	Relieff Weight	Rank	Feature	Relieff Weight
1	Cor Qwy	0.3047		Range Ax	0.1327		SD Qw	0.0932		Variance Mx	0.0791
2	Cor Axz	0.2791		SD Mx	0.1292		75% Mx	0.0932		Entropy Gy	0.0787
3	Cor Qwz	0.2743		95% Ax	0.1246		SD Az	0.0931		50% Az	0.0782
4	Cor Qwx	0.2609		RMS Az	0.1241		Mean Ey	0.0926		75% Az	0.0779
5	Cor Qxy	0.2603	40	Entropy Mx	0.1234		Entropy My	0.0920	110	IQR Ay	0.0778
6	Cor Exy	0.2466		5% Ax	0.1201		25% My	0.0919		Entropy Gx	0.0743
7	Cor Myz	0.2406		Range My	0.1195		25% Ay	0.0908		95% Ay	0.0737
8	Cor Qxz	0.2401		5% Ey	0.1177		IQ Range Qz	0.0907		95% Mz	0.0727
9	Cor Ayz	0.2219		95% Ey	0.1172		50% Ey	0.0905		5% Mz	0.0725
10	Cor Axy	0.2140		IQR Mx	0.1138	80	IQR My	0.0902		75% Mz	0.0722
	Cor Mxy	0.2079		SD My	0.1127		50% My	0.0900		50% Mz	0.0719
	Cor Qyz	0.2035		RMS Ey	0.1124		SD Ay	0.0900		Mean Mz	0.0718
	Cor Mxz	0.1987		Cor Gxy	0.1124		50% Mx	0.0896		25% Mz	0.0714
	RMS Ax	0.1928		Mean Qy	0.1121		Mean Az	0.0888		Entropy Az	0.0712
	Cor Exz	0.1830	50	Variance Ax	0.1111		25% Az	0.0886	120	5% Mxyz	0.0706
	Entropy Ax	0.1775		RMS Qz	0.1097		Range Mz	0.0883		95% Mxyz	0.0703
	Cor Gyz	0.1706		25% Ey	0.1095		Mean Freq Ax	0.0882		25% Mxyz	0.0690
	75% Ax	0.1668		95% My	0.1078		SD Mz	0.0880		Variance My	0.0677
	Cor Eyz	0.1667		RMS Mx	0.1075		IQR Mxyz	0.0879		75% Mxyz	0.0675
20	Mean Qx	0.1630		Range Mxyz	0.1049	90	Mean Ay	0.0877		Mean Mxyz	0.0675
	SD Ax	0.1575		Range Qz	0.1042		IQ Range Qw	0.0873		RMS Mxyz	0.0674
	Range Ey	0.1573		SD Mxyz	0.1026		IQR Az	0.0869		RMS Qx	0.0664
	Cor Gxz	0.1563		75% Ey	0.1009		Entropy Ey	0.0867		50% Mxyz	0.0657
	RMS Ay	0.1559		5% Ay	0.1006		50% Ay	0.0866		RMS Axyz	0.0638
	Entropy Ex	0.1548	60	SD Qz	0.1002		Mean Qw	0.0863	130	Variance Mxyz	0.0635
	SD Ey	0.1517		5% Mx	0.0999		RMS Qw	0.0861		SD Qy	0.0633
	IQR Ax	0.1481		Range Qw	0.0981		Range Az	0.0860		Range Ez	0.0611
	RMS My	0.1433		25% Mx	0.0980		25% Axyz	0.0859		Variance Qz	0.0593
	Mean Qz	0.1392		75% My	0.0980		RMS Mz	0.0839		RMS Gy	0.0592
30	25% Ax	0.1377		Variance Ey	0.0973	100	95% Az	0.0833		Range Ay	0.0590
	50% Ax	0.1360		95% Mx	0.0962		75% Ay	0.0821		Mean Freq Gx	0.0579
	Range Mx	0.1355		RMS Qy	0.0951		Entropy Gz	0.0817		Mean Freq Gy	0.0576
	Entropy Ez	0.1352		Mean My	0.0942		IQR Mz	0.0809		Range Qy	0.0574
	IQR Ey	0.1341		Mean Mx	0.0938		Mean Axyz	0.0800		IQ Range Qx	0.0571
	Mean Ax	0.1339	70	5% My	0.0938		5% Az	0.0791	140	Entropy Mxyz	0.0566

Rank	Feature	Relieff Weight	Rank	Feature	Relieff Weight	Rank	Feature	Relieff Weight	Rank	Feature	Relieff Weight
	Variance Qw	0.0559		SD Axyz	0.0367		75% Gz	0.0232		Energy Gz	0.0113
	50% Axyz	0.0558		75% Gxyz	0.0364		50% Gx	0.0229		Max Freq Axyz	0.0109
	Variance Az	0.0548		5% Ex	0.0360		SD Gz	0.0228		Energy Axyz	0.0108
	IQ Range Qy	0.0546		25% Gxyz	0.0360		SD Gx	0.0226		Variance Ez	0.0103
	RMS Ez	0.0545	180	IQR Axyz	0.0345		50% Gz	0.0226	250	Energy Gxyz	0.0099
	Variance Ay	0.0532		75% Ex	0.0342		5% Gz	0.0220		Variance Gy	0.0090
	SD Qx	0.0517		Energy My	0.0342		SD Gxyz	0.0217		Kurtosis Ay	0.0084
	Mean Freq My	0.0502		Mean Ex	0.0323		IQR Ex	0.0217		Max Freq Mx	0.0084
	5% Ez	0.0502		Mean Freq Az	0.0317		5% Gxyz	0.0214		Kurtosis Mxyz	0.0083
150	Mean Freq Gz	0.0488		25% Ex	0.0316	220	50% Gy	0.0214		Max Freq Mxyz	0.0081
	Variance Mz	0.0487		RMS Gxyz	0.0315		75% Gy	0.0213		Kurtosis My	0.0081
	50% Gxyz	0.0471		SD Gy	0.0312		Mean Freq Ay	0.0212		Kurtosis Mz	0.0081
	RMS Ex	0.0468		50% Ex	0.0310		95% Gxyz	0.0204		Variance Ex	0.0079
	5% Axyz	0.0463		IQR Gx	0.0309		95% Gx	0.0203		Kurtosis Gy	0.0079
	Range Qx	0.0457	190	Mean Gx	0.0301		5% Gy	0.0193	260	Kurtosis Gx	0.0073
	Entropy Axyz	0.0441		75% Gx	0.0299		Range Gxyz	0.0191		Kurtosis Ey	0.0071
	Max Freq Gz	0.0432		SD Ex	0.0293		Entropy Gxyz	0.0187		Energy Mxyz	0.0070
	Range Ex	0.0432		Energy Az	0.0281		Range Gy	0.0183		Energy Gx	0.0068
	95% Ez	0.0432		Range Axyz	0.0281		Energy Ay	0.0177		Variance Gxyz	0.0062
160	RMS Gx	0.0430		Mean Gz	0.0280	230	Max Freq My	0.0176		Kurtosis Ax	0.0058
	Max Freq Gy	0.0421		Max Freq Gxyz	0.0276		Energy Gy	0.0167		Kurtosis Gz	0.0057
	IQR Gy	0.0420		95% Axyz	0.0269		Max Freq Ax	0.0164		Energy Mx	0.0055
	25% Ez	0.0412		IQR Gxyz	0.0266		Max Freq Az	0.0160		Variance Gx	0.0052
	IQR Gz	0.0411		Mean Freq Mxyz	0.0262		Variance Qx	0.0160		Kurtosis Gxyz	0.0051
	Entropy Ay	0.0410	200	Max Freq Gx	0.0261		Range Gx	0.0157	270	Kurtosis Axyz	0.0050
	SD Ez	0.0408		95% Gy	0.0259		Energy Mz	0.0154		Kurtosis Ex	0.0049
	RMS Gz	0.0406		Variance Qy	0.0259		Energy Ax	0.0142		Variance Gz	0.0043
	Mean Gxyz	0.0398		Mean Gy	0.0248		Variance Axyz	0.0140		Kurtosis Ez	0.0022
	Mean Ez	0.0396		IQR Ez	0.0245		Range Gz	0.0137			
170	75% Ez	0.0392		Mean Freq Mx	0.0241	240	Kurtosis Az	0.0136			
	Mean Freq Gxyz	0.0390		5% Gx	0.0239		Max Freq Ay	0.0135			
	50% Ez	0.0388		25% Gx	0.0239		Mean Freq Axyz	0.0128			
	Entropy Mz	0.0374		25% Gy	0.0236		Max Freq Mz	0.0125			
	75% Axyz	0.0367		25% Gz	0.0233		Kurtosis Mx	0.0125			
	95% Ex	0.0367	210	Mean Freq Mz	0.0232		95% Gz	0.0115			

Table C2. All features ranked from highest to lowest based on ReliefF algorithm weight.

Appendix D

Contents

D.1 MATLAB code **174**
D.2 Interview topic guide **189**

D.1 MATLAB code

PRE-PROCESSING

```
% Import elapsed time and 3 axis data from csv file as numeric matrix

Acc=csvread('AAcc.csv', 1, 2);
Gyro=csvread('AGyro.csv', 1, 2);
Mag=csvread('AMag.csv', 1, 2);



%% Reconstruct accelerometer time stamp and create consistently sampled
dataset

Acc(:,1)=round(Acc(:,1), 2); % round timestamps to nearest 1ms

Acc(2:3:end,1)=Acc(2:3:end,1)+0.01; % replace every second time stamp n
for n+0.01

Acc(3:3:end,1)=Acc(3:3:end,1)+0.02; % replace every third time stamp n
for n+0.02

[~,idx] = unique(Acc(:,1)); % remove repeating values
Acc = Acc(idx,:);

At=Acc(:,1); % define new raw time series data after reconstructing
time stamp and removing repeating values

Axyz=Acc(:,2:4);

AT=0:0.01:At(end); % define new time array according to required
sampling rate (100Hz) and file length

ATl=int32((At(end)*100)+1); % define number of data points required to
match new time array

Ats = timeseries(Axyz, At,'name', 'Acc'); % create time series with new
raw data (this enables interpolation of missing data while retaining
all original data points)

Ats = setinterpmethod(Ats,'zoh'); % set zero-order hold interpolation
method

Ats1 = resample(Ats,[AT]); % resample keeping all original values

AccelerometerUF=getdatasamples(Ats1,[1:ATl]); % extract values from
time series

%low pass filtering - 6th order Butterworth bidirectional filter

Ad1 = designfilt('lowpassiir','FilterOrder',6, ...
```

```
    'HalfPowerFrequency',0.10,'DesignMethod','butter');
Accelerometer1 = filtfilt(Ad1,AccelerometerUF);



%% Reconstruct gyroscope time stamp and create consistently sampled
dataset

Gyro(:,1)=round(Gyro(:,1), 2); % round timestamps to nearest 1ms

Gyro(2:3:end,1)=Gyro(2:3:end,1)+0.01; % replace every second time stamp
n for n+0.01

Gyro(3:3:end,1)=Gyro(3:3:end,1)+0.02; % replace every third time stamp
n for n+0.02

[~,idx] = unique(Gyro(:,1)); % remove repeating values
Gyro = Gyro(idx,:);

Gt=Gyro(:,1); % define new raw time series data after reconstructing
time stamp and removing repeating values

Gxyz=Gyro(:,2:4);

GT=0:0.01:Gt(end); % define new time array according to required
sampling rate (100Hz) and file length

GTl=int32((Gt(end)*100)+1); % define number of data points required to
match new time array

Gts = timeseries(Gxyz, Gt,'name', 'Gyro'); % create time series with
new raw data (this enables interpolation of missing data while
retaining all original data points)

Gts = setinterpmethod(Gts,'zoh'); % set zero-order hold interpolation
method

Gts1 = resample(Gts,[GT]); % resample keeping all original values

Gyroscope1=getdatasamples(Gts1,[1:GTl]); % extract values from time
series



% % Correct gyroscope offset (only used if stationary for first 5
seconds)
%
% Gyroscopex=Gyroscope1(:,1)- mean(Gyroscope1(1:500,1));
% Gyroscopey=Gyroscope1(:,2)- mean(Gyroscope1(1:500,2));
% Gyroscopez=Gyroscope1(:,3)- mean(Gyroscope1(1:500,3));
%
% Gyroscope1=[Gyroscopex Gyroscopey Gyroscopez];



%% Reconstruct magnetometer time stamp and create consistently sampled
dataset

Mag(:,1)=round(Mag(:,1), 2); % round timestamps to nearest 1ms
```

```
Mt=Mag(:,1); % define new raw time series data after reconstructing
time stamp and removing repeating values

Mxyz=Mag(:,2:4);

MT=0:0.01:Mt(end); % define new time array according to required
sampling rate (100Hz) and file length

MTl=int32((Mt(end)*100)+1); % define number of data points required to
match new time array

Mts = timeseries(Mxyz, Mt,'name', 'Mag'); % create time series with new
raw data (this enables interpolation of missing data while retaining
all original data points)

Mts = setinterpmethod(Mts,'zoh'); % set zero-order hold interpolation
method

Mts1 = resample(Mts,[MT]); % resample keeping all original values

Magnetometer1=getdatasamples(Mts1, [1:MTl]); % extract values from time
series



% Create series matched 9 axis data set at 100Hz

timex=round(Acc(end,1))-1; % create new time series that falls within
length of all 3 data sets
times=(timex*100)+1;
time=(0:0.01:timex)';



% define 3 axis data for common time stamp

Accelerometer=Accelerometer1(1:times,:);
Gyroscope=Gyroscope1(1:times,:);
Magnetometer=Magnetometer1(1:times,:);

% calculate sensor vector magnitudes

VAcc=vecnorm(Accelerometer, 2, 2);
VGyro=vecnorm(Gyroscope, 2, 2);
VMag=vecnorm(Magnetometer, 2, 2);
```

SENSOR FUSION (MADGWICK FILTER)

```
addpath('quaternion_library');      % include quaternion library

AHRS = MadgwickAHRS('SamplePeriod', 1/100, 'Beta', 0.1); % adjustable
parameters for filter

quaternion = zeros(length(timex), 4);
for t = 1:length(time)
```

```
    AHRS.Update(Gyroscope(t,:) * (pi/180), Accelerometer(t,:),
    Magnetometer(t,:));
    quaternion(t, :) = AHRS.Quaternion;
end

euler = quatern2euler(quaternConj(quaternion)) * (180/pi); % convert
quaternions to Euler angles
```

SEGMENTATION

```
% Using a peak analysis to identify when gyro vector stream is closest
to zero

% MinPeakProminence and MinPeakDistance are adjustable variables

[pks, locs] = findpeaks(-VGyro(:,1), 'MinPeakProminence', 30, ...
     'MinPeakDistance', 80);
hold off

indices = find(pks<-40); % remove locs with peaks < 40

locs(indices) = [];


% Define start/end points for all segments

locsEnd=locs(end)+1;
locs1=[locs(1:end)];
locs2=[locs(2:end);locsEnd];
```

FEATURE CALCULATION AND CLASSIFICATION

```
load('Classifier02_01.mat');
load('BlankTable.mat');

Tend=length(locs1);     % define number of segments

AT = zeros(1,119);      % create empty feature table (for classifier)
ST = zeros(1,6);       % create empty score table

Index=(1:Tend)';

%LocationsTable=[0 0 0; Index locs1 locs2];

Ts = 0.01;  % Ts needs defining for SPARC function
```

```
%% Calculation of features for all segments and creation of feature
table

for n=1:Tend

   % quaternion (19 features)

    Mqw=mean(quaternion(locs1(n):locs2(n),1));

    Mqx=mean(quaternion(locs1(n):locs2(n),2));

    Mqy=mean(quaternion(locs1(n):locs2(n),3));

    Mqz=mean(quaternion(locs1(n):locs2(n),4));


    RMSqw=rms(quaternion(locs1(n):locs2(n),1));

    RMSqy=rms(quaternion(locs1(n):locs2(n),3));

    RMSqz=rms(quaternion(locs1(n):locs2(n),4));

    Sqw=std(quaternion(locs1(n):locs2(n),1));

    Sqz=std(quaternion(locs1(n):locs2(n),4));

    Rqw=range(quaternion(locs1(n):locs2(n),1));

    Rqz=range(quaternion(locs1(n):locs2(n),4));

    IQRqw=iqr(quaternion(locs1(n):locs2(n),1));

    IQRqz=iqr(quaternion(locs1(n):locs2(n),4));

    Cqwx1=corrcoef(quaternion(locs1(n):locs2(n),1),
quaternion(locs1(n):locs2(n),2));

    Cqwx=Cqwx1(1,2);

    Cqwy1=corrcoef(quaternion(locs1(n):locs2(n),1),
quaternion(locs1(n):locs2(n),3));

    Cqwy=Cqwy1(1,2);

    Cqwz1=corrcoef(quaternion(locs1(n):locs2(n),1),
quaternion(locs1(n):locs2(n),4));

    Cqwz=Cqwz1(1,2);

    Cqxy1=corrcoef(quaternion(locs1(n):locs2(n),2),
quaternion(locs1(n):locs2(n),3));

    Cqxy=Cqxy1(1,2);

    Cqxz1=corrcoef(quaternion(locs1(n):locs2(n),2),
quaternion(locs1(n):locs2(n),4));

    Cqxz=Cqxz1(1,2);
```

```
    Cqyz1=corrcoef(quaternion(locs1(n):locs2(n),3),
quaternion(locs1(n):locs2(n),4));

    Cqyz=Cqyz1(1,2);

    % euler (17 features)

    Mey=mean(euler(locs1(n):locs2(n),2));

    RMSey=rms(euler(locs1(n):locs2(n),2));

    Sey=std(euler(locs1(n):locs2(n),2));

    Vey=var(euler(locs1(n):locs2(n),2));

    Rey=range(euler(locs1(n):locs2(n),2));

    IQRey=iqr(euler(locs1(n):locs2(n),2));

    pc5ey=prctile(euler(locs1(n):locs2(n),2),5);

    pc25ey=prctile(euler(locs1(n):locs2(n),2),25);

    pc50ey=prctile(euler(locs1(n):locs2(n),2),50);

    pc75ey=prctile(euler(locs1(n):locs2(n),2),75);

    pc95ey=prctile(euler(locs1(n):locs2(n),2),95);

    Cexy1=corrcoef(euler(locs1(n):locs2(n),1),
euler(locs1(n):locs2(n),2));

    Cexy=Cexy1(1,2);

    Cexz1=corrcoef(euler(locs1(n):locs2(n),1),
euler(locs1(n):locs2(n),3));

    Cexz=Cexz1(1,2);

    Ceyz1=corrcoef(euler(locs1(n):locs2(n),2),
euler(locs1(n):locs2(n),3));

    Ceyz=Ceyz1(1,2);

    ex_norm =
euler(locs1(n):locs2(n),1)/max(abs(euler(locs1(n):locs2(n),1)));
    Entex = entropy(ex_norm);

    ey_norm =
euler(locs1(n):locs2(n),2)/max(abs(euler(locs1(n):locs2(n),2)));
    Entey = entropy(ey_norm);

    ez_norm =
euler(locs1(n):locs2(n),3)/max(abs(euler(locs1(n):locs2(n),3)));
    Entez = entropy(ez_norm);



    % accelerometer (36 features)
```

```
Max=mean(Accelerometer(locs1(n):locs2(n),1));
May=mean(Accelerometer(locs1(n):locs2(n),2));
Maz=mean(Accelerometer(locs1(n):locs2(n),3));
Maxyz=mean(VAcc(locs1(n):locs2(n)));
RMSax=rms(Accelerometer(locs1(n):locs2(n),1));
RMSay=rms(Accelerometer(locs1(n):locs2(n),2));
RMSaz=rms(Accelerometer(locs1(n):locs2(n),3));
Sax=std(Accelerometer(locs1(n):locs2(n),1));
Say=std(Accelerometer(locs1(n):locs2(n),2));
Saz=std(Accelerometer(locs1(n):locs2(n),3));
Vax=var(Accelerometer(locs1(n):locs2(n),1));
Rax=range(Accelerometer(locs1(n):locs2(n),1));
Raz=range(Accelerometer(locs1(n):locs2(n),3));
IQRax=iqr(Accelerometer(locs1(n):locs2(n),1));
IQRay=iqr(Accelerometer(locs1(n):locs2(n),2));
IQRaz=iqr(Accelerometer(locs1(n):locs2(n),3));
pc5ax=prctile(Accelerometer(locs1(n):locs2(n),1),5);
pc5ay=prctile(Accelerometer(locs1(n):locs2(n),2),5);
pc5az=prctile(Accelerometer(locs1(n):locs2(n),3),5);
pc25ax=prctile(Accelerometer(locs1(n):locs2(n),1),25);
pc25ay=prctile(Accelerometer(locs1(n):locs2(n),2),25);
pc25az=prctile(Accelerometer(locs1(n):locs2(n),3),25);
pc25axyz=prctile(VAcc(locs1(n):locs2(n)),25);
pc50ax=prctile(Accelerometer(locs1(n):locs2(n),1),50);
pc50ay=prctile(Accelerometer(locs1(n):locs2(n),2),50);
pc50az=prctile(Accelerometer(locs1(n):locs2(n),3),50);
pc75ax=prctile(Accelerometer(locs1(n):locs2(n),1),75);
pc75ay=prctile(Accelerometer(locs1(n):locs2(n),2),75);
pc75az=prctile(Accelerometer(locs1(n):locs2(n),3),75);
pc95ax=prctile(Accelerometer(locs1(n):locs2(n),1),95);
```

```
    pc95ay=prctile(Accelerometer(locs1(n):locs2(n),2),95);

    pc95az=prctile(Accelerometer(locs1(n):locs2(n),3),95);


    Caxy1=corrcoef(Accelerometer(locs1(n):locs2(n),1),
Accelerometer(locs1(n):locs2(n),2));

    Caxy=Caxy1(1,2);

    Caxz1=corrcoef(Accelerometer(locs1(n):locs2(n),1),
Accelerometer(locs1(n):locs2(n),3));

    Caxz=Caxz1(1,2);

    Cayz1=corrcoef(Accelerometer(locs1(n):locs2(n),2),
Accelerometer(locs1(n):locs2(n),3));

    Cayz=Cayz1(1,2);


    ax_norm =
Accelerometer(locs1(n):locs2(n),1)/max(abs(Accelerometer(locs1(n):locs2
(n),1)));
    Entax = entropy(ax_norm);

    az_norm =
Accelerometer(locs1(n):locs2(n),3)/max(abs(Accelerometer(locs1(n):locs2
(n),3)));
    Entaz = entropy(az_norm);


    % gyroscope (6 features)

    Cgxy1=corrcoef(Gyroscope(locs1(n):locs2(n),1),
Gyroscope(locs1(n):locs2(n),2));

    Cgxy=Cgxy1(1,2);

    Cgxz1=corrcoef(Gyroscope(locs1(n):locs2(n),1),
Gyroscope(locs1(n):locs2(n),3));

    Cgxz=Cgxz1(1,2);

    Cgyz1=corrcoef(Gyroscope(locs1(n):locs2(n),2),
Gyroscope(locs1(n):locs2(n),3));

    Cgyz=Cgyz1(1,2);


    gx_norm =
Gyroscope(locs1(n):locs2(n),1)/max(abs(Gyroscope(locs1(n):locs2(n),1)))
;
    Entgx = entropy(gx_norm);

    gy_norm =
Gyroscope(locs1(n):locs2(n),2)/max(abs(Gyroscope(locs1(n):locs2(n),2)))
;
    Entgy = entropy(gy_norm);
```

```
    gz_norm =
Gyroscope(locs1(n):locs2(n),3)/max(abs(Gyroscope(locs1(n):locs2(n),3)))
;
    Entgz = entropy(gz_norm);

    % magnetometer (39 features)

    Mmx=mean(Magnetometer(locs1(n):locs2(n),1));

    Mmy=mean(Magnetometer(locs1(n):locs2(n),2));

    Mmz=mean(Magnetometer(locs1(n):locs2(n),3));

    RMSmx=rms(Magnetometer(locs1(n):locs2(n),1));

    RMSmy=rms(Magnetometer(locs1(n):locs2(n),2));

    RMSmz=rms(Magnetometer(locs1(n):locs2(n),3));

    Smx=std(Magnetometer(locs1(n):locs2(n),1));

    Smy=std(Magnetometer(locs1(n):locs2(n),2));

    Smz=std(Magnetometer(locs1(n):locs2(n),3));

    Smxyz=std(VMag(locs1(n):locs2(n)));

    Vmx=var(Magnetometer(locs1(n):locs2(n),1));

    Rmx=range(Magnetometer(locs1(n):locs2(n),1));

    Rmy=range(Magnetometer(locs1(n):locs2(n),2));

    Rmz=range(Magnetometer(locs1(n):locs2(n),3));

    Rmxyz=range(VMag(locs1(n):locs2(n)));

    IQRmx=iqr(Magnetometer(locs1(n):locs2(n),1));

    IQRmy=iqr(Magnetometer(locs1(n):locs2(n),2));

    IQRmz=iqr(Magnetometer(locs1(n):locs2(n),3));

    IQRmxyz=iqr(VMag(locs1(n):locs2(n)));

    pc5mx=prctile(Magnetometer(locs1(n):locs2(n),1),5);

    pc5my=prctile(Magnetometer(locs1(n):locs2(n),2),5);

    pc5mz=prctile(Magnetometer(locs1(n):locs2(n),3),5);

    pc5mxyz=prctile(VMag(locs1(n):locs2(n)),5);

    pc25mx=prctile(Magnetometer(locs1(n):locs2(n),1),25);

    pc25my=prctile(Magnetometer(locs1(n):locs2(n),2),25);

    pc25mz=prctile(Magnetometer(locs1(n):locs2(n),3),25);

    pc50mx=prctile(Magnetometer(locs1(n):locs2(n),1),50);
```

```
    pc50my=prctile(Magnetometer(locs1(n):locs2(n),2),50);

    pc50mz=prctile(Magnetometer(locs1(n):locs2(n),3),50);

    pc75mx=prctile(Magnetometer(locs1(n):locs2(n),1),75);

    pc75my=prctile(Magnetometer(locs1(n):locs2(n),2),75);

    pc75mz=prctile(Magnetometer(locs1(n):locs2(n),3),75);

    pc95mx=prctile(Magnetometer(locs1(n):locs2(n),1),95);

    pc95my=prctile(Magnetometer(locs1(n):locs2(n),2),95);

    pc95mz=prctile(Magnetometer(locs1(n):locs2(n),3),95);

    Cmxy1=corrcoef(Magnetometer(locs1(n):locs2(n),1),
Magnetometer(locs1(n):locs2(n),2));

    Cmxy=Cmxy1(1,2);

    Cmxz1=corrcoef(Magnetometer(locs1(n):locs2(n),1),
Magnetometer(locs1(n):locs2(n),3));

    Cmxz=Cmxz1(1,2);

    Cmyz1=corrcoef(Magnetometer(locs1(n):locs2(n),2),
Magnetometer(locs1(n):locs2(n),3));

    Cmyz=Cmyz1(1,2);


    mx_norm =
Magnetometer(locs1(n):locs2(n),1)/max(abs(Magnetometer(locs1(n):locs2(n
),1)));
    Entmx = entropy(mx_norm);

    my_norm =
Magnetometer(locs1(n):locs2(n),2)/max(abs(Magnetometer(locs1(n):locs2(n
),2)));
    Entmy = entropy(my_norm);



    % Calculate scores for all segments

    % SPARC metric to quantify smoothness

    Sparcx = SpectralArcLength(Gyroscope(locs1(n):locs2(n),1), Ts);

    Sparcy = SpectralArcLength(Gyroscope(locs1(n):locs2(n),2), Ts);

    Sparcz = SpectralArcLength(Gyroscope(locs1(n):locs2(n),3), Ts);

    Sparc = (Sparcx + Sparcy + Sparcz)/3;

    % Power metric

    VAccRange = range(VAcc(locs1(n):locs2(n)));
```

```
    VGyroRange = range(VGyro(locs1(n):locs2(n)));

    Power = VAccRange*VGyroRange;


    % Dominant axis movement range for each exercise


    qinv=quatinv(quaternion(locs1(n),:));

    qproduct=quatmultiply(qinv, quaternion(locs2(n),:));

    RadAngle=2*acos(qproduct(1));

    ROMQ=rad2deg(RadAngle);


    % Update features table (for classifier)

    NewAT=[Mqw Mqx Mqy Mqz RMSqw RMSqy RMSqz Sqw Sqz Rqw Rqz IQRqw
IQRqz Cqwx Cqwy Cqwz Cqxy Cqxz Cqyz Mey RMSey Sey Vey Rey IQRey pc5ey
pc25ey pc50ey pc75ey pc95ey Cexy Cexz Ceyz Entex Entey Entez Max May
Maz Maxyz RMSax RMSay RMSaz Sax Say Saz Vax Rax Raz IQRax IQRay IQRaz
pc5ax pc5ay pc5az pc25ax pc25ay pc25az pc25axyz pc50ax pc50ay pc50az
pc75ax pc75ay pc75az pc95ax pc95ay pc95az Caxy Caxz Cayz Entax Entaz
Cgxy Cgxz Cgyz Entgx Entgy Entgz Mmx Mmy Mmz RMSmx RMSmy RMSmz Smx Smy
Smz Smxyz Vmx Rmx Rmy Rmz Rmxyz IQRmx IQRmy IQRmz IQRmxyz pc5mx pc5my
pc5mz pc5mxyz pc25mx pc25my pc25mz pc50mx pc50my pc50mz pc75mx pc75my
pc75mz pc95mx pc95my pc95mz Cmxy Cmxz Cmyz Entmx Entmy];

    AT=[AT;NewAT];

    % Update Score table

    NewST = [Sparc Power ROMWS ROMWP ROMR ROMQ];

    ST = [ST;NewST];

    n=n+1;

end


% create activity (features) table for testing

ActivityTable=array2table(AT); % table format for classifier

ActivityTable1=[BlankTable; ActivityTable]; % merge with BlankTable to
ensure columns have correct labels

ActivityTable1=ActivityTable(2:end,:); % remove blank row

% classifier (Random Forests) output

yfit = Classifier02_01.predictFcn(ActivityTable1)
```

SCORE CALCULATION

```
ST1 = ST;

% create score table

ST = ST1(2:end,:); % remove blank row

SPARC = ST(:,1);
Power = ST(:,2);
ROMWS = ST(:,3);
ROMWP = ST(:,4);
ROMR = ST(:,5);
ROMQ = ST(:,6);

% Wall slide metrics

% for loop to identify which rows of score table ("ST") correspond to
wall
% slide activity

WSCount=[0];

for n=1:Tend

    WSOutput = isequal(yfit(n), 'WALL SLIDE');

    WSCount= [WSCount WSOutput];

    n=n+1;
end

WSCount=WSCount(2:end);

WSCount=WSCount';

WSTable = [WSCount SPARC Power ROMWS ROMQ];

% If statement to compile new score table only for wall slide segments

WSTable1=[0, 0, 0, 0 0];

for n=1:Tend

    if  WSTable((n),1)==1;

        NWSRow = WSTable((n),:);

        WSTable1=[WSTable1; NWSRow];

    end

    n=n+1;
end

WSTable1=WSTable1(2:end,:);

WallSlideCount = size(WSTable1, 1)/2; %% CHANGED
```

```
WSSPARC = rmoutliers(WSTable1(:,2));
WSMeanSPARC = mean(WSSPARC);
WSSDSPARC = std(WSSPARC);

WSPower = rmoutliers(WSTable1(:,3));
WSMeanPower = mean(WSPower);
WSSDPower = std(WSPower);

WSROM = rmoutliers(WSTable1(:,4));
WSMeanROM = mean(WSROM);
WSSDROM = std(WSROM);

WSROMQ = rmoutliers(WSTable1(:,5));
WSMeanROMQ = mean(WSROMQ);
WSSDROMQ = std(WSROMQ);


% Wall press metrics

% for loop to identify which rows of score table ("ST") correspond to
wall
% press activity

WPCount=[0];

for n=1:Tend

    WPOutput = isequal(yfit(n), 'WALL PRESS');

    WPCount= [WPCount WPOutput];

    n=n+1;
end

WPCount=WPCount(2:end);

WPCount=WPCount';

WPTable = [WPCount SPARC Power ROMWP ROMQ];


% If statement to compile new score table only for wall press segments

WPTable1=[0, 0, 0, 0 0];

for n=1:Tend

    if  WPTable((n),1)==1;

        NWPRow = WPTable((n),:);

        WPTable1=[WPTable1; NWPRow];

    end

    n=n+1;
end

WPTable1=WPTable1(2:end,:);
```

```
WallPressCount = size(WPTable1, 1)/2; %% CHANGED

WPSPARC = rmoutliers(WPTable1(:,2));
WPMeanSPARC = mean(WPSPARC);
WPSDSPARC = std(WPSPARC);

WPPower = rmoutliers(WPTable1(:,3));
WPMeanPower = mean(WPPower);
WPSDPower = std(WPPower);

WPROM = rmoutliers(WPTable1(:,4));
WPMeanROM = mean(WPROM);
WPSDROM = std(WPROM);

WPROMQ = rmoutliers(WPTable1(:,5));
WPMeanROMQ = mean(WPROMQ);
WPSDROMQ = std(WPROMQ);

% Rotation metrics

% for loop to identify which rows of score table ("ST") correspond to
rotation activity

RotCount=[0];

for n=1:Tend

    ROutput = isequal(yfit(n), 'ROTATION');

    RotCount= [RotCount ROutput];

    n=n+1;
end

RotCount=RotCount(2:end);

RotCount=RotCount';

RotTable = [RotCount SPARC Power ROMR ROMQ];

% If statement to compile new score table only for rotation segments

RotTable1=[0, 0, 0, 0 0];

for n=1:Tend

    if  RotTable((n),1)==1;

        NRotRow = RotTable((n),:);

        RotTable1=[RotTable1; NRotRow];

    end

    n=n+1;
end

RotTable1=RotTable1(2:end,:);

RotationCount = size(RotTable1, 1)/2; %% CHANGED
```

```
RotSPARC = rmoutliers(RotTable1(:,2));
RotMeanSPARC = mean(RotSPARC);
RotSDSPARC = std(RotSPARC);

RotPower = rmoutliers(RotTable1(:,3));
RotMeanPower = mean(RotPower);
RotSDPower = std(RotPower);

RotROM = rmoutliers(RotTable1(:,4));
RotMeanROM = mean(RotROM);
RotSDROM = std(RotROM);

RotROMQ = rmoutliers(RotTable1(:,5));
RotMeanROMQ = mean(RotROMQ);
RotSDROMQ = std(RotROMQ);

Columns = {'Count';'SPARC';'Power';'ROM';'ROMQ'};
Rows = {'Wall Slide', 'Wall Press', 'Rotation'};
Count = [WallSlideCount; WallPressCount; RotationCount];
SPARC = [WSMeanSPARC; WPMeanSPARC; RotMeanSPARC];
Power = [WSMeanPower; WPMeanPower; RotMeanPower];
ROM = [WSMeanROM; WPMeanROM; RotMeanROM];
ROMQ = [WSMeanROMQ; WPMeanROMQ; RotMeanROMQ];
T = table(Count, SPARC, Power, ROM, ROMQ, 'VariableNames',Columns,
'RowNames', Rows)
```

D.2 Interview topic guide

The following questions formed the main structure of the interview that took place during the follow-up review. Further questions were asked depending on the subject's initial response and previous comments.

How did you find using this system?

Example of follow-up questions:

Did you have any difficulty using the system by yourself?

How long did it take you to get comfortable with it?

Was anything about the process particularly challenging?

Would you be happy to continue using it?

Do you think using this system affected your rehabilitation in any way?

Example of follow-up questions:

How did recording your exercises make you feel?

Were you more or less inclined to do your exercises?

Why do you think that is?

Did you experience any problems whilst using the application and were there any tasks that you found particularly difficult?

Example of follow-up questions:

How often did this happen?

Do you remember what you were doing or what was on the application when this happened?

How did this affect you and your exercises?

What did you think about the application and how information was presented?

Example of follow-up questions:

Was it useful to see how much exercise you had done?

Do you think it was accurate?

Were the progress graphs helpful?

Based on your experience of using this system, can you think any way we can improve it?

Example of follow-up questions:

How would you like that data presented to you?

Do you have any other thoughts about this system, or about the use of sensor technology in rehabilitation?

References

1. Urwin M, Symmons D, Allison T, Brammah T, Busby H, Roxby M, et al. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. Ann Rheum Dis. 1998;57(11):649-55.
2. Mitchell C, Adebajo A, Hay E, Carr A. Shoulder pain: diagnosis and management in primary care. BMJ. 2005;331(7525):1124-8.
3. Virta L, Joranger P, Brox JI, Eriksson R. Costs of shoulder pain and resource use in primary health care: a cost-of-illness study in Sweden. BMC Musculoskelet Disord. 2012;13:17.
4. Jones T, Carr AJ, Beard D, Linton MJ, Rooshenas L, Donovan J, et al. Longitudinal study of use and cost of subacromial decompression surgery: the need for effective evaluation of surgical procedures to prevent overtreatment and wasted resources. BMJ Open. 2019;9(8):e030229.
5. Mroz TM, Carlini AR, Archer KR, Wegener ST, Hoolachan JI, Stiers W, et al. Frequency and cost of claims by injury type from a state workers' compensation fund from 1998 through 2008. Arch Phys Med Rehabil. 2014;95(6):1048-54 e6.
6. Littlewood C, Ashton J, Chance-Larsen K, May S, Sturrock B. Exercise for rotator cuff tendinopathy: a systematic review. Physiotherapy. 2012;98(2):101-9.
7. Beard DJ, Rees JL, Cook JA, Rombach I, Cooper C, Merritt N, et al. Arthroscopic subacromial decompression for subacromial shoulder pain (CSAW): a multicentre, pragmatic, parallel group, placebo-controlled, three-group, randomised surgical trial. Lancet. 2018;391(10118):329-38.
8. Paavola M, Malmivaara A, Taimela S, Kanto K, Inkinen J, Kalske J, et al. Subacromial decompression versus diagnostic arthroscopy for shoulder impingement: randomised, placebo surgery controlled clinical trial. BMJ. 2018;362:k2860.
9. Heron SR, Woby SR, Thompson DP. Comparison of three types of exercise in the treatment of rotator cuff tendinopathy/shoulder impingement syndrome: A randomized controlled trial. Physiotherapy. 2017;103(2):167-73.
10. Shire AR, Staehr TAB, Overby JB, Bastholm Dahl M, Sandell Jacobsen J, Hoyrup Christiansen D. Specific or general exercise strategy for subacromial impingement syndrome-does it matter? A systematic literature review and meta analysis. BMC Musculoskelet Disord. 2017;18(1):158.
11. Physotherapy CSo. Streched to the Limit. An audit of physiotherapy services in England. Full Report.; 2012.
12. Bollen JC, Dean SG, Siegert RJ, Howe TE, Goodwin VA. A systematic review of measures of self-reported adherence to unsupervised home-based rehabilitation exercise programmes, and their psychometric properties. BMJ Open. 2014;4(6):e005044.
13. Holden MA, Haywood KL, Potia TA, Gee M, McLean S. Recommendations for exercise adherence measures in musculoskeletal settings: a systematic review and consensus meeting (protocol). Syst Rev. 2014;3:10.
14. Bassett S. The assessment of patient adherence to physiotherapy rehabilitation. New Zealand Journal of Physiotherapy. 2003;31:60-6.
15. Wang Q, Markopoulos P, Yu B, Chen W, Timmermans A. Interactive wearable systems for upper body rehabilitation: a systematic review. J Neuroeng Rehabil. 2017;14(1):20.

16. Brennan DM, Mawson S, Brownsell S. Telerehabilitation: enabling the remote delivery of healthcare, rehabilitation, and self management. Stud Health Technol Inform. 2009;145:231-48.
17. Khusainov R, Azzi D, Achumba IE, Bersch SD. Real-time human ambulation, activity, and physiological monitoring: taxonomy of issues, techniques, applications, challenges and limitations. Sensors (Basel). 2013;13(10):12852-902.
18. Papi E, Murtagh GM, McGregor AH. Wearable technologies in osteoarthritis: a qualitative study of clinicians' preferences. BMJ Open. 2016;6(1):e009544.
19. Hochstenbach-Waelen A, Seelen HA. Embracing change: practical and theoretical considerations for successful implementation of technology assisting upper limb training in stroke. J Neuroeng Rehabil. 2012;9:52.
20. Demain S, Burridge J, Ellis-Hill C, Hughes AM, Yardley L, Tedesco-Triccas L, et al. Assistive technologies after stroke: self-management or fending for yourself? A focus group study. BMC health services research. 2013;13:334.
21. Veeger HE, van der Helm FC. Shoulder function: the perfect compromise between mobility and stability. J Biomech. 2007;40(10):2119-29.
22. Halder AM, Itoi E, An KN. Anatomy and biomechanics of the shoulder. Orthop Clin North Am. 2000;31(2):159-76.
23. Escamilla RF, Yamashiro K, Paulos L, Andrews JR. Shoulder muscle activity and function in common shoulder rehabilitation exercises. Sports Med. 2009;39(8):663-85.
24. Garving C, Jakob S, Bauer I, Nadjar R, Brunner UH. Impingement Syndrome of the Shoulder. Dtsch Arztebl Int. 2017;114(45):765-76.
25. Neer CS, 2nd. Anterior acromioplasty for the chronic impingement syndrome in the shoulder. 1972. J Bone Joint Surg Am. 2005;87(6):1399.
26. Neer CS, 2nd. Anterior acromioplasty for the chronic impingement syndrome in the shoulder: a preliminary report. J Bone Joint Surg Am. 1972;54(1):41-50.
27. Judge A, Murphy RJ, Maxwell R, Arden NK, Carr AJ. Temporal trends and geographical variation in the use of subacromial decompression and rotator cuff repair of the shoulder in England. Bone Joint J. 2014;96-B(1):70-4.
28. Ludewig PM, Lawrence RL, Braman JP. What's in a name? Using movement system diagnoses versus pathoanatomic diagnoses. J Orthop Sports Phys Ther. 2013;43(5):280-3.
29. Papadonikolakis A, McKenna M, Warme W, Martin BI, Matsen FA, 3rd. Published evidence relevant to the diagnosis of impingement syndrome of the shoulder. J Bone Joint Surg Am. 2011;93(19):1827-32.
30. Seitz AL, McClure PW, Finucane S, Boardman ND, 3rd, Michener LA. Mechanisms of rotator cuff tendinopathy: intrinsic, extrinsic, or both? Clin Biomech (Bristol, Avon). 2011;26(1):1-12.
31. Michener LA, Walsworth MK, Burnet EN. Effectiveness of rehabilitation for patients with subacromial impingement syndrome: a systematic review. J Hand Ther. 2004;17(2):152-64.
32. Ludewig PM, Braman JP. Shoulder impingement: biomechanical considerations in rehabilitation. Man Ther. 2011;16(1):33-9.
33. Littlewood C, May S, S. W. A review of systematic reviews of the effectiveness of conservative interventions for rotator cuff tendinopathy. Shoulder Elbow. 2013;5:151-67.
34. Hanratty CE, McVeigh JG, Kerr DP, Basford JR, Finch MB, Pendleton A, et al. The effectiveness of physiotherapy exercises in subacromial impingement syndrome: a systematic review and meta-analysis. Semin Arthritis Rheum. 2012;42(3):297-316.
35. Gebremariam L, Hay EM, van der Sande R, Rinkel WD, Koes BW, Huisstede BM. Subacromial impingement syndrome--effectiveness of physiotherapy and manual therapy. Br J Sports Med. 2014;48(16):1202-8.

36. Kromer TO, Tautenhahn UG, de Bie RA, Staal JB, Bastiaenen CH. Effects of physiotherapy in patients with shoulder impingement syndrome: a systematic review of the literature. J Rehabil Med. 2009;41(11):870-80.
37. Page MJ, Green S, McBain B, Surace SJ, Deitch J, Lyttle N, et al. Manual therapy and exercise for rotator cuff disease. Cochrane Database Syst Rev. 2016(6):CD012224.
38. Steuri R, Sattelmayer M, Elsig S, Kolly C, Tal A, Taeymans J, et al. Effectiveness of conservative interventions including exercise, manual therapy and medical management in adults with shoulder impingement: a systematic review and meta-analysis of RCTs. Br J Sports Med. 2017;51(18):1340-7.
39. Bennell K, Wee E, Coburn S, Green S, Harris A, Staples M, et al. Efficacy of standardised manual therapy and home exercise programme for chronic rotator cuff disease: randomised placebo controlled trial. BMJ. 2010;340:c2756.
40. Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. Man Ther. 2010;15(3):220-8.
41. Nicolson PJA, Hinman RS, Wrigley TV, Stratford PW, Bennell KL. Self-reported Home Exercise Adherence: A Validity and Reliability Study Using Concealed Accelerometers. J Orthop Sports Phys Ther. 2018;48(12):943-50.
42. Yuen HK, Wang E, Holthaus K, Vogtle LK, Sword D, Breland HL, et al. Self-reported versus objectively assessed exercise adherence. Am J Occup Ther. 2013;67(4):484-9.
43. Moseley GL. Do training diaries affect and reflect adherence to home programs? Arthritis Rheum. 2006;55(4):662-4.
44. Roetenberg D, Slycke PJ, Veltink PH. Ambulatory position and orientation tracking fusing magnetic and inertial sensing. IEEE Trans Biomed Eng. 2007;54(5):883-90.
45. O'Donovan KJ, Kamnik R, O'Keeffe DT, Lyons GM. An inertial and magnetic sensor based technique for joint angle measurement. J Biomech. 2007;40(12):2604-11.
46. Roetenberg D, Luinge HJ, Baten CT, Veltink PH. Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation. IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society. 2005;13(3):395-405.
47. Schiefer C, Ellegast RP, Hermanns I, Kraus T, Ochsmann E, Larue C, et al. Optimization of inertial sensor-based motion capturing for magnetically distorted field applications. Journal of biomechanical engineering. 2014;136(12):121008.
48. Luinge HJ, Veltink PH. Measuring orientation of human body segments using miniature gyroscopes and accelerometers. Med Biol Eng Comput. 2005;43(2):273-82.
49. Diaz EM, Ahmed DB, Kaiser S. A Review of Indoor Localization Methods Based on Inertial Sensors. Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation: Academic Press; 2019. p. 311-33.
50. Garcia-Alsina J, Garcia Almazan C, Moranta Mesquida J, Pleguezuelos Cobo E. Angular position, range of motion and velocity of arm elevation: a study of consistency of performance. Clin Biomech (Bristol, Avon). 2005;20(9):932-8.
51. Klopcar N, Lenarcic J. Bilateral and unilateral shoulder girdle kinematics during humeral elevation. Clin Biomech (Bristol, Avon). 2006;21 Suppl 1:S20-6.
52. Leggin BG, Neuman RM, Iannotti JP, Williams GR, Thompson EC. Intrarater and interrater reliability of three isometric dynamometers in assessing shoulder strength. J Shoulder Elbow Surg. 1996;5(1):18-24.
53. van Andel CJ, Wolterbeek N, Doorenbosch CA, Veeger DH, Harlaar J. Complete 3D kinematics of upper extremity functional tasks. Gait Posture. 2008;27(1):120-7.

54. Veeger HE, Magermans DJ, Nagels J, Chadwick EK, van der Helm FC. A kinematical analysis of the shoulder after arthroplasty during a hair combing task. Clin Biomech (Bristol, Avon). 2006;21 Suppl 1:S39-44.
55. Cutti AG, Chadwick EK. Shoulder biomechanics and the success of translational research. Med Biol Eng Comput. 2014;52(3):205-10.
56. de los Reyes-Guzman A, Dimbwadyo-Terrer I, Trincado-Alonso F, Monasterio-Huelin F, Torricelli D, Gil-Agudo A. Quantitative assessment based on kinematic measures of functional impairments during upper extremity movements: A review. Clin Biomech (Bristol, Avon). 2014;29(7):719-27.
57. https://www.wearable-technologies.com [Available from: https://www.wearable-technologies.com.
58. Tunca C, Pehlivan N, Ak N, Arnrich B, Salur G, Ersoy C. Inertial Sensor-Based Robust Gait Analysis in Non-Hospital Settings for Neurological Disorders. Sensors (Basel). 2017;17(4).
59. De Baets L, van der Straaten R, Matheve T, Timmermans A. Shoulder assessment according to the international classification of functioning by means of inertial sensor technologies: A systematic review. Gait Posture. 2017;57:278-94.
60. Schall MC, Jr., Fethke NB, Chen H, Oyama S, Douphrate DI. Accuracy and repeatability of an inertial measurement unit system for field-based occupational studies. Ergonomics. 2016;59(4):591-602.
61. El-Gohary M, McNames J. Shoulder and elbow joint angle tracking with inertial sensors. IEEE Trans Biomed Eng. 2012;59(9):2635-41.
62. Fantozzi S, Giovanardi A, Magalhaes FA, Di Michele R, Cortesi M, Gatta G. Assessment of three-dimensional joint kinematics of the upper limb during simulated swimming using wearable inertial-magnetic measurement units. J Sports Sci. 2016;34(11):1073-80.
63. Bouvier B, Duprey S, Claudon L, Dumas R, Savescu A. Upper Limb Kinematics Using Inertial and Magnetic Sensors: Comparison of Sensor-to-Segment Calibrations. Sensors (Basel). 2015;15(8):18813-33.
64. Picerno P, Viero V, Donati M, Triossi T, Tancredi V, Melchiorri G. Ambulatory assessment of shoulder abduction strength curve using a single wearable inertial sensor. J Rehabil Res Dev. 2015;52(2):171-80.
65. Coley B, Jolles BM, Farron A, Aminian K. Arm position during daily activity. Gait Posture. 2008;28(4):581-7.
66. Cutti AG, Paolini G, Troncossi M, Cappello A, Davalli A. Soft tissue artefact assessment in humeral axial rotation. Gait Posture. 2005;21(3):341-9.
67. Parel I, Cutti AG, Fiumana G, Porcellini G, Verni G, Accardo AP. Ambulatory measurement of the scapulohumeral rhythm: intra- and inter-operator agreement of a protocol based on inertial and magnetic sensors. Gait Posture. 2012;35(4):636-40.
68. Coley B, Jolles BM, Farron A, Bourgeois A, Nussbaumer F, Pichonnaz C, et al. Outcome evaluation in shoulder surgery using 3D kinematics sensors. Gait Posture. 2007;25(4):523-32.
69. Korver RJ, Senden R, Heyligers IC, Grimm B. Objective outcome evaluation using inertial sensors in subacromial impingement syndrome: a five-year follow-up study. Physiol Meas. 2014;35(4):677-86.
70. Pichonnaz C, Duc C, Gleeson N, Ancey C, Jaccard H, Lecureux E, et al. Measurement properties of the smartphone-based B-B Score in current shoulder pathologies. Sensors (Basel). 2015;15(10):26801-17.
71. Korver RJ, Heyligers IC, Samijo SK, Grimm B. Inertia based functional scoring of the shoulder in clinical practice. Physiol Meas. 2014;35(2):167-76.

72. Kennedy DM, Stratford PW, Wessel J, Gollish JD, Penney D. Assessing stability and change of four performance measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty. BMC Musculoskelet Disord. 2005;6:3.
73. Freter SH, Fruchter N. Relationship between timed 'up and go' and gait time in an elderly orthopaedic rehabilitation population. Clin Rehabil. 2000;14(1):96-101.
74. Balasubramanian S, Melendez-Calderon A, Burdet E. A robust and sensitive metric for quantifying movement smoothness. IEEE Trans Biomed Eng. 2012;59(8):2126-36.
75. Carpinella I, Cattaneo D, Abuarqub S, Ferrarin M. Robot-based rehabilitation of the upper limbs in multiple sclerosis: feasibility and preliminary results. J Rehabil Med. 2009;41(12):966-70.
76. Tresilian JR, Stelmach GE, Adler CH. Stability of reach-to-grasp movement patterns in Parkinson's disease. Brain : a journal of neurology. 1997;120 (Pt 11):2093-111.
77. Rand MK, Stelmach GE, Bloedel JR. Movement accuracy constraints in Parkinson's disease patients. Neuropsychologia. 2000;38(2):203-12.
78. Gulde P, Hermsdorfer J. Smoothness Metrics in Complex Movement Tasks. Front Neurol. 2018;9:615.
79. Beck Y, Herman T, Brozgol M, Giladi N, Mirelman A, Hausdorff JM. SPARC: a new approach to quantifying gait smoothness in patients with Parkinson's disease. J Neuroeng Rehabil. 2018;15(1):49.
80. Newman CJ, Bruchez R, Roches S, Jequier Gygax M, Duc C, Dadashi F, et al. Measuring upper limb function in children with hemiparesis with 3D inertial sensors. Childs Nerv Syst. 2017;33(12):2159-68.
81. Fukaya T, Mutsuzaki H, Nakano W, Mori K. Smoothness of the knee joint movement during the stance phase in patients with severe knee osteoarthritis. Asia Pac J Sports Med Arthrosc Rehabil Technol. 2018;14:1-5.
82. Chan M, Esteve D, Fourniols JY, Escriba C, Campo E. Smart wearable systems: current status and future challenges. Artif Intell Med. 2012;56(3):137-56.
83. Lee SS, Choi ST, Choi SI. Classification of Gait Type Based on Deep Learning Using Various Sensors with Smart Insole. Sensors (Basel). 2019;19(8).
84. Attal F, Mohammed S, Dedabrishvili M, Chamroukhi F, Oukhellou L, Amirat Y. Physical Human Activity Recognition Using Wearable Sensors. Sensors (Basel). 2015;15(12):31314-38.
85. Taylor PE, Almeida GJ, Kanade T, Hodgins JK. Classifying human motion quality for knee osteoarthritis using accelerometers. Conf Proc IEEE Eng Med Biol Soc. 2010;2010:339-43.
86. O'Reilly M, Caulfield B, Ward T, Johnston W, Doherty C. Wearable Inertial Sensor Systems for Lower Limb Exercise Detection and Evaluation: A Systematic Review. Sports Med. 2018;48(5):1221-46.
87. Pan J, Chung H, J. H. Intelligent Shoulder Joint Home-Based Self-Rehabilitation Monitoring System. International Journal of Smart Home. 2013;7(5):395-404.
88. Burns DM, Leung N, Hardisty M, Whyne CM, Henry P, McLachlin S. Shoulder physiotherapy exercise recognition: machine learning the inertial signals from a smartwatch. Physiol Meas. 2018;39(7):075007.
89. Carroll N. Key Success Factors for Smart and Connected Health Software Solutions. Computer. 2016;49(11):22-8.
90. Carroll N, Kennedy C, Richardson I. Connected Community Healthcare Ecosystem (CCHE) for managing long-term conditions. Gerontechnology. 2016;14(2):64-77.
91. Kairy D, Lehoux P, Vincent C, Visintin M. A systematic review of clinical outcomes, clinical process, healthcare utilization and costs associated with telerehabilitation. Disabil Rehabil. 2009;31(6):427-47.

92. Ayoade M, Baillie L. A novel knee rehabilitation system for the home. Proceedings of the 32nd annual ACM conference on Human factors in computing systems. 2014;2521-30.
93. Park JH, Lee SH, Ko DS. The Effects of the Nintendo Wii Exercise Program on Chronic Work-related Low Back Pain in Industrial Workers. J Phys Ther Sci. 2013;25(8):985-8.
94. Lin DH, Lin YF, Chai HM, Han YC, Jan MH. Comparison of proprioceptive functions between computerized proprioception facilitation exercise and closed kinetic chain exercise in patients with knee osteoarthritis. Clin Rheumatol. 2007;26(4):520-8.
95. Kim SS, Min WK, Kim JH, Lee BH. The Effects of VR-based Wii Fit Yoga on Physical Function in Middle-aged Female LBP Patients. J Phys Ther Sci. 2014;26(4):549-52.
96. Huang MC, Lee SH, C. YS, Chan RC, Rizzo A, Xu W. Intelligent Frozen Shoulder Rehabilitation. IEEE Intelligent Systems. 2014;29:22-8.
97. Hershko E, Tauber C, Carmeli E. Biofeedback versus physiotherapy in patients with partial weight-bearing. Am J Orthop (Belle Mead NJ). 2008;37(5):E92-6.
98. Piqueras M, Marco E, Coll M, Escalada F, Ballester A, Cinca C, et al. Effectiveness of an interactive virtual telerehabilitation system in patients after total knee arthoplasty: a randomized controlled trial. J Rehabil Med. 2013;45(4):392-6.
99. Grimm B, Bolink S. Evaluating physical function and activity in the elderly patient using wearable motion sensors. EFORT open reviews. 2016;1(5):112-20.
100. Lebel K, Boissy P, Nguyen H, Duval C. Inertial measurement systems for segments and joints kinematics assessment: towards an understanding of the variations in sensors accuracy. Biomedical engineering online. 2017;16(1):56.
101. Lebel K, Boissy P, Hamel M, Duval C. Inertial measures of motion for clinical biomechanics: comparative assessment of accuracy under controlled conditions - effect of velocity. PLoS One. 2013;8(11):e79945.
102. Ricci L, Taffoni F, Formica D. On the Orientation Error of IMU: Investigating Static and Dynamic Accuracy Targeting Human Motion. PLoS One. 2016;11(9):e0161940.
103. Madgwick SO, Harrison AJ, Vaidyanathan A. Estimation of IMU and MARG orientation using a gradient descent algorithm. IEEE International Conference on Rehabilitation Robotics : [proceedings]. 2011;2011:5975346.
104. Kalman R. A new approach to linear filtering and prediction problems. Journal of Basic Engineering. 1960;82:35-45.
105. Feng K, Li J, Zhang X, Shen C, Bi Y, Zheng T, et al. A New Quaternion-Based Kalman Filter for Real-Time Attitude Estimation Using the Two-Step Geometrically-Intuitive Correction Algorithm. Sensors (Basel). 2017;17(9).
106. Mahony R, Hamel T, Pflimlin J-M. Nonlinear Complementary Filters on the Special Orthogonal Group. IEEE Transactions on Automatic Control, Institute of Electrical and Electronics
Engineers. 2008;53(5):1203-17.
107. Diebel J. Representing attitude: Euler angles, unit quaternions, and rotation vectors. Matrix. 2006;58(15-16).
108. Brodie MA, Walmsley A, Page W. The static accuracy and calibration of inertial measurement units for 3D orientation. Computer methods in biomechanics and biomedical engineering. 2008;11(6):641-8.
109. Cutti AG, Giovanardi A, Rocchi L, Davalli A. A simple test to assess the static and dynamic accuracy of an inertial sensors system for human movement analysis. Conf Proc IEEE Eng Med Biol Soc. 2006;1:5912-5.
110. Picerno P, Cereatti A, Cappozzo A. A spot check for assessing static orientation consistency of inertial and magnetic sensing units. Gait Posture. 2011;33(3):373-8.

111. Godwin A, Agnew M, Stevenson J. Accuracy of inertial motion sensors in static, quasistatic, and complex dynamic motion. Journal of biomechanical engineering. 2009;131(11):114501.
112. Bergamini E, Ligorio G, Summa A, Vannozzi G, Cappozzo A, Sabatini AM. Estimating orientation using magnetic and inertial sensors and different sensor fusion approaches: accuracy assessment in manual and locomotion tasks. Sensors (Basel). 2014;14(10):18625-49.
113. Lebel K, Boissy P, Hamel M, Duval C. Inertial measures of motion for clinical biomechanics: comparative assessment of accuracy under controlled conditions - changes in accuracy over time. PLoS One. 2015;10(3):e0118361.
114. Technologies X-I. [Available from: http://x-io.co.uk/open-source-imu-and-ahrs-algorithms/.
115. Smith SW. The scientist and engineer's guide to digital signal processing. San Diego, California: California Technical Publishing; 1997.
116. Godfrey A, Conway R, Meagher D, G OL. Direct measurement of human movement by accelerometry. Medical engineering & physics. 2008;30(10):1364-86.
117. Ayub S, Bahraminisaab A, Honary B, editors. A sensor fusion method for smart phone orientation estimation. 13th Annual Post Graduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting; 2012; Liverpool, UK.
118. Mann KA, Werner FW, Palmer AK. Frequency spectrum analysis of wrist motion for activities of daily living. Journal of orthopaedic research : official publication of the Orthopaedic Research Society. 1989;7(2):304-6.
119. Feng K, Li J, Zhang X, Shen C, Bi Y, Zheng T, et al. Correction: A New Quaternion-Based Kalman Filter for Real-Time Attitude Estimation Using the Two-Step Geometrically-Intuitive Correction Algorithm. Sensors 2017, 17, 2146. Sensors (Basel). 2017;17(11).
120. Zhang ZQ, Yang GZ. Micromagnetometer calibration for accurate orientation estimation. IEEE Trans Biomed Eng. 2015;62(2):553-60.
121. Yang D, You Z, Li B, Duan W, Yuan B. Complete Tri-Axis Magnetometer Calibration with a Gyro Auxiliary. Sensors (Basel). 2017;17(6).
122. Gates DH, Walters LS, Cowley J, Wilken JM, Resnik L. Range of Motion Requirements for Upper-Limb Activities of Daily Living. Am J Occup Ther. 2016;70(1):7001350010p1-p10.
123. Ho CY, Sole G, Munn J. The effectiveness of manual therapy in the management of musculoskeletal disorders of the shoulder: a systematic review. Man Ther. 2009;14(5):463-74.
124. Riddle DL, Rothstein JM, Lamb RL. Goniometric reliability in a clinical setting. Shoulder measurements. Phys Ther. 1987;67(5):668-73.
125. Chang FM, Seidl AJ, Muthusamy K, Meininger AK, Carollo JJ. Effectiveness of instrumented gait analysis in children with cerebral palsy--comparison of outcomes. J Pediatr Orthop. 2006;26(5):612-6.
126. Valevicius AM, Jun PY, Hebert JS, Vette AH. Use of optical motion capture for the analysis of normative upper body kinematics during functional upper limb tasks: A systematic review. J Electromyogr Kinesiol. 2018;40:1-15.
127. Simon SR. Quantification of human motion: gait analysis-benefits and limitations to its application to clinical problems. J Biomech. 2004;37(12):1869-80.
128. Parel I, Jaspers E, L DEB, Amoresano A, Cutti AG. Motion analysis of the shoulder in adults: kinematics and electromyography for the clinical practice. European journal of physical and rehabilitation medicine. 2016;52(4):575-82.
129. Roldan-Jimenez C, Cuesta-Vargas AI. Studying upper-limb kinematics using inertial sensors: a cross-sectional study. BMC Res Notes. 2015;8:532.

130. Wu G, van der Helm FC, Veeger HE, Makhsous M, Van Roy P, Anglin C, et al. ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion--Part II: shoulder, elbow, wrist and hand. J Biomech. 2005;38(5):981-92.
131. [Available from: https://docs.vicon.com/display/Nexus210.
132. Mcgraw KO, Wong SP. Forming Inferences About Some Intraclass Correlation Coefficients. Psychological Methods. 1996;1(1):30-46.
133. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155-63.
134. Zhou H, Hu H. Human motion tracking for rehabilitation - a survey. Biomed Signal Proccess Control. 2008;3(1):1-18.
135. Hegedus EJ, Goode A, Campbell S, Morin A, Tamaddoni M, Moorman CT, 3rd, et al. Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. Br J Sports Med. 2008;42(2):80-92; discussion
136. Terwee CB, van der Slikke RM, van Lummel RC, Benink RJ, Meijers WG, de Vet HC. Self-reported physical functioning was more influenced by pain than performance-based physical functioning in knee-osteoarthritis patients. Journal of clinical epidemiology. 2006;59(7):724-31.
137. Stratford PW, Kennedy DM, Woodhouse LJ. Performance measures provide assessments of pain and function in people with advanced osteoarthritis of the hip or knee. Phys Ther. 2006;86(11):1489-96.
138. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. BJA Education. 2017;17(4):137-44.
139. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med. 2006;119(2):166 e7-16.
140. Davenport TE, Stevens SR, Baroni K, Van Ness JM, Snell CR. Reliability and validity of Short Form 36 Version 2 to measure health perceptions in a sub-group of individuals with fatigue. Disabil Rehabil. 2011;33(25-26):2596-604.
141. Balasubramanian S, Melendez-Calderon A, Roby-Brami A, Burdet E. On the analysis of movement smoothness. J Neuroeng Rehabil. 2015;12:112.
142. Balasubramanian S. [Available from: https://github.com/siva82kb/SPARC.
143. Pichonnaz C, Lecureux E, Bassin JP, Duc C, Farron A, Aminian K, et al. Enhancing clinically-relevant shoulder function assessment using only essential movements. Physiol Meas. 2015;36(3):547-60.
144. Ekeberg OM, Bautz-Holter E, Tveita EK, Keller A, Juel NG, Brox JI. Agreement, reliability and validity in 3 shoulder questionnaires in patients with rotator cuff disease. BMC Musculoskelet Disord. 2008;9:68.
145. Magermans DJ, Chadwick EK, Veeger HE, van der Helm FC. Requirements for upper extremity motions during activities of daily living. Clin Biomech (Bristol, Avon). 2005;20(6):591-9.
146. van der Windt DA, Koes BW, de Jong BA, Bouter LM. Shoulder disorders in general practice: incidence, patient characteristics, and management. Ann Rheum Dis. 1995;54(12):959-64.
147. Yamaguchi K, Ditsios K, Middleton WD, Hildebolt CF, Galatz LM, Teefey SA. The demographic and morphological features of rotator cuff disease. A comparison of asymptomatic and symptomatic shoulders. J Bone Joint Surg Am. 2006;88(8):1699-704.
148. Bassett SF. The assessment of patient adherence to physiotherapy rehabilitation. NZ Journal of Physiotherapy. 2003;31:60-2.

149. Room J, Hannink E, Dawes H, Barker K. What interventions are used to improve exercise adherence in older people and what behavioural techniques are they based on? A systematic review. BMJ Open. 2017;7(12):e019221.
150. Lin HC, Chiang SY, Lee K, Kan YC. An activity recognition model using inertial sensor nodes in a wireless sensor network for frozen shoulder rehabilitation exercises. Sensors (Basel). 2015;15(1):2181-204.
151. Giggins OM, Sweeney KT, Caulfield B. Rehabilitation exercise assessment using inertial sensors: a cross-sectional analytical study. J Neuroeng Rehabil. 2014;11:158.
152. Verbrugghe J, Knippenberg E, Palmaers S, Matheve T, Smeets W, Feys P, et al. Motion detection supported exercise therapy in musculoskeletal disorders: a systematic review. European journal of physical and rehabilitation medicine. 2018;54(4):591-604.
153. Palazzo C, Klinger E, Dorner V, Kadri A, Thierry O, Boumenir Y, et al. Barriers to home-based exercise program adherence with chronic low back pain: Patient expectations regarding new technologies. Annals of physical and rehabilitation medicine. 2016;59(2):107-13.
154. https://www.bess.org.uk/index.php/public-area/shpi-videos [
155. Gonzalez S, Sedano J, Villar JR, Corchado E, Herrero., Baruque B. Features and models for human activity recognition. Neurocomputing. 2015;167:52-60.
156. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003;3:1157-82.
157. Kohavi R, John GH. Wrappers for feature subset selection. Artificial Intelligence. 1997;97(1-2):273-324.
158. Kononenko I. Estimating attributes: Analysis ad extensions of RELIEF. Lect Notes Comput Sci 1994(784):171-82.
159. Lopez C, Tucker S, Salameh T, Tucker C. An unsupervised machine learning method for discovering patient clusters based on genetic signatures. J Biomed Inform. 2018;85:30-9.
160. Janidarmian M, Roshan Fekr A, Radecka K, Zilic Z. A Comprehensive Analysis on Wearable Acceleration Sensors in Human Activity Recognition. Sensors (Basel). 2017;17(3).
161. Schiele B, Stikic M, Van Laerhoven K. Exploring semi-supervised and active learning for activity recognition. 2008 12th IEEE International Symposium on Wearable Computers(ISWC); Pittaburgh, PA, USA2008. p. 81-8.
162. Feng Z, Mo L, Li M. A Random Forest-based ensemble method for activity recognition. Conf Proc IEEE Eng Med Biol Soc. 2015;2015:5074-7.
163. Banos O, Galvez JM, Damas M, Pomares H, Rojas I. Window size impact in human activity recognition. Sensors (Basel). 2014;14(4):6474-99.
164. Dougherty ER, Hua J, Sima C. Performance of feature selection methods. Current genomics. 2009;10(6):365-74.
165. Allen FR, Ambikairajah E, Lovell NH, Celler BG. Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models. Physiol Meas. 2006;27(10):935-51.
166. Laguna O, Olaya A, Borrajo D. A Dynamic Sliding Window Approach for Activity Recognition. Lecture Notes in Computer Science. 2011;6787.
167. Zapata BC, Fernandez-Aleman JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst. 2015;39(2):1.
168. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res. 2017;19(6):e232.
169. Bergmann JH, McGregor AH. Body-worn sensor design: what do patients and clinicians want? Ann Biomed Eng. 2011;39(9):2299-312.

170. Peake JM, Kerr G, Sullivan JP. A Critical Review of Consumer Wearables, Mobile Applications, and Equipment for Providing Biofeedback, Monitoring Stress, and Sleep in Physically Active Populations. Front Physiol. 2018;9:743.
171. O'Reilly MA, Slevin P, Ward T, Caulfield B. A Wearable Sensor-Based Exercise Biofeedback System: Mixed Methods Evaluation of Formulift. JMIR Mhealth Uhealth. 2018;6(1):e33.
172. Jordan P, Thomas B, McClelland I, Weerdmeester B. Usability Evaluation in Industry. London, UK: Taylor and Francis; 1996.
173. Bangor A, Kortum P, Miller L. An Empirical Evaluation of the Sytem Usability Scale. Int J Hum-Comput Interact. 2008(24):574-94.
174. Bangor A, Kortum P, Miller L. Determining what individual SUS scores mean: Adding an adjective rating scale. J Usability Stud. 2009(4):114-23.
175. J. S. Measuring usability with the System Usability Scale (SUS) Userfocus.co.uk 2016 [Available from: https://www.useerfocus.co.uk/articles/measuring-usability-with-the-SUS.html.
176. Argent R, Slevin P, Bevilacqua A, Neligan M, Daly A, Caulfield B. Wearable Sensor-Based Exercise Biofeedback for Orthopaedic Rehabilitation: A Mixed Methods User Evaluation of a Prototype System. Sensors (Basel). 2019;19(2).
177. Chughtai M, Shah NV, Sultan AA, Solow M, Tiberi JV, Mehran N, et al. The role of prehabilitation with a telerehabilitation system prior to total knee arthroplasty. Ann Transl Med. 2019;7(4):68.
178. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. Contemp Clin Trials Commun. 2018;11:156-64.
179. Reiman MP, Manske RC. The assessment of function: How is it measured? A clinical perspective. J Man Manip Ther. 2011;19(2):91-9.
180. Forner-Cordero A, Mateu-Arce M, Forner-Cordero I, Alcantara E, Moreno JC, Pons JL. Study of the motion artefacts of skin-mounted inertial sensors under different attachment conditions. Physiol Meas. 2008;29(4):N21-31.
181. Smith S, Cano S, Browne J. Patient reported outcome measurement: drawbacks of existing methods. BMJ. 2019;364:l844.
182. Booker S, Alfahad N, Scott M, Gooding B, Wallace WA. Use of scoring systems for assessing and reporting the outcome results from shoulder surgery and arthroplasty. World J Orthop. 2015;6(2):244-51.
183. Chau T. A review of analytical techniques for gait data. Part 1: Fuzzy, statistical and fractal methods. Gait Posture. 2001;13(1):49-66.
184. Goto M, Takedani H, Haga N, Kubota M, Ishiyama M, Ito S, et al. Self-monitoring has potential for home exercise programmes in patients with haemophilia. Haemophilia. 2014;20(2):e121-7.