Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

Håvard Pettersen Reierstad

Exploring Frequency, Variability and Propulsion Patterns in Manual Wheelchair Propulsion

Master's thesis in Cybernetics and Robotics
Supervisor: Damiano Varagnolo
Co-supervisor: Julia Kathrin Baumgart & Roya Doshmanziari
June 2023

NTNU
Norwegian University of
Science and Technology

Håvard Pettersen Reierstad

Exploring Frequency, Variability and Propulsion Patterns in Manual Wheelchair Propulsion

Master's thesis in Cybernetics and Robotics
Supervisor: Damiano Varagnolo
Co-supervisor: Julia Kathrin Baumgart & Roya Doshmanziari
June 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of
Science and Technology

Preface

This thesis concludes my five-year Master's program in Cybernetics and Robotics with a specialization in Biomedical Cybernetics at the Norwegian University of Science and Technology (NTNU) in Trondheim. The thesis is a continuation of a specialization project that was written in the fall of 2022, that explored features related to the frequency and variability of manual wheelchair propulsion. The specialization project examined data from 12 participants without disability, and this thesis extends this to include data from 20 participants without disability and 15 manual wheelchair users. Data from 20 manual wheelchair users has been collected, but only the data from the first 15 was available at the time of this thesis.

The data used in this thesis was gathered as part of a cross-sectional research project involving various departments from NTNU, among others the Department of Neuroscience and Movement Science and the Department of Engineering Cybernetics. The dataset consists of various motion data and physiological data from manual wheelchair propulsion, and is further described in Section 4 along with the equipment used to collect it. The specific motion data used in this thesis consists of accelerometer signals from inertial measurement units and motion trajectories captured using Qualisys Track Manager. The specific physiological data investigated are the energy expenditure and gross efficiency, which was measured using gold standard techniques in a lab setting. The analysis was conducted using MATLAB.

I would like to thank my supervisors Damiano Varagnolo, Roya Doshmanziari and Julia Kathrin Baumgart for their guidance and insights. Working on the same project over the course of a year has been both challenging and interesting, and I appreciate their support along the way. Finally, I would like to thank the people at Toppidrettssenteret at Granåsen for providing me with helpful insights into the field of biomechanics, as well as insights into how the data collection was conducted. Coming from a field of cybernetics, it was incredibly useful to exchange information with people with different academic backgrounds.

Abstract

Technological advances in recent years have facilitated the use of low-cost multi-sensor devices for activity and fitness tracking. However, most of these devices are primarily designed for people without disabilities and are thus not suitable for manual wheelchair users. This population has a higher prevalence of obesity and cardiovascular disease, highlighting the need for tools specifically adapted to their needs.

This thesis aims to contribute to the development of technologies tailored for wheelchair users. The primary focus is to investigate the correlation between specific features derived from manual wheelchair propulsion and physiological values, with the goal of utilizing these features in a biofeedback tool. Two main objectives were pursued; firstly, to investigate features related to the frequency and variability in manual wheelchair propulsion movements and their relation to energy expenditure and gross efficiency, and secondly, to demonstrate the extraction of these features using equipment that is available outside of a lab setting, specifically inertial measurement units.

The results highlight a correlation between the frequency of movement and energy expenditure. Specifically, the data collected and analyzed suggest that the number of cycles within a given time window is positively correlated with energy expenditure. This feature can be reliably extracted from inertial measurement units.

Additionally, this thesis presents novel methods for examining variability in propulsion patterns, utilizing tools from statistical shape analysis and time series analysis. The results suggest that variability in propulsion patterns is influenced by the speed and incline of the movement. However, it remains unclear whether propulsion techniques with low or high variability should be recommended to wheelchair users. Extraction of variability features relies on motion trajectories, which were able to reproduced with moderate success from acceleration signals using a long short-term memory neural network.

Future work should focus on exploring the implications of the variability-based features in relation to physiological values. Advancements in the extraction and analysis techniques may offer insights into optimizing wheelchair propulsion techniques and developing personalized interventions for manual wheelchair users.

Sammendrag

Teknologiske fremskritt i de siste årene har gjort det mulig å ta i bruk ulike sensorer for aktivitets- og treningsmåling. De fleste av disse enhetene er hovedsakelig beregnet for personer uten funksjonsnedsettelser, og er dermed ikke egnet for manuelle rullestolbrukere. Denne gruppen har høyere forekomst av fedme og hjerte- og karsykdommer, og det er derfor behov for enheter som er tilpasset deres behov.

Denne avhandlingen har som mål å bidra til utviklingen av teknologi som er tilpasset rullestolbrukere. Hovedfokuset er å undersøke sammenhengen mellom spesifikke egenskaper ved manuell rullestolfremdrift og fysiologiske verdier, med mål om å bruke disse egenskapene i en aktivitetsmåler som kan gi tilbakemelding til brukeren. Avhandlingen bestod av to hovedmål; for det første å undersøke egenskaper relatert til frekvens og variasjon i armbevegelser under manuell rullestolfremdrift, og deres forhold til energiforbruk og bruttoeffektivitet, og for det andre å demonstrere utvinning av disse egenskapene ved hjelp av utstyr som er tilgjengelig utenfor et laboratoriemiljø, spesifikt inertielle måleenheter.

Resultatene fremhever en sammenheng mellom bevegelsesfrekvens og energiforbruk. Spesifikt antyder de innsamlede og analyserte dataene at antallet hånd-sykluser innenfor et gitt tidsvindu har en positiv korrelasjon med energiforbruket. Denne egenskapen kan pålitelig ekstraheres fra inertielle måleenheter.

I tillegg presenterer denne avhandlingen nye metoder for å undersøke variasjon i fremdriftsmønstre ved hjelp av verktøy fra statistisk formanalyse og tidsrekkeanalyse. Resultatene tyder på at variasjonen i fremdriftsmønstre påvirkes av hastigheten og stigningen i bevegelsen. Det er imidlertid uklart om fremdriftsteknikker med lav eller høy variasjon bør anbefales for rullestolbrukere. Utvinning av variasjonsegenskaper er avhengig av bevegelsesbaner, som kunne gjenskapes med moderat suksess fra akselerasjonssignaler ved hjelp av et nevralt nettverk.

Fremtidig arbeid bør fokusere på å utforske konsekvensene av variasjonsbaserte egenskaper i forhold til fysiologiske verdier. Forbedringer i utvinnings- og analysemetodene kan gi innsikt i optimalisering av fremdriftsteknikker og utvikling av personlig tilpassede tiltak for manuelle rullestolbrukere.

Contents

List of Figures vi

List of Tables viii

1 Introduction 1

1.1 Background . . . 1

1.2 Research Goals . . . 1

2 Literature Review 3

2.1 Propulsion Patterns in Manual Wheelchair Propulsion . . . 3

2.2 Classification of Propulsion Patterns . . . 4

2.3 Variability in Manual Wheelchair Propulsion . . . 4

2.4 Motion Tracking Using Inertial Measurement Units . . . 4

3 Theory 6

3.1 Physiological Values . . . 6

3.1.1 Energy Expenditure . . . 6

3.1.2 Power Output . . . 6

3.1.3 Gross Efficiency . . . 6

3.2 Principal Component Analysis . . . 6

3.3 Similarity Measures . . . 7

3.3.1 Procrustes Analysis . . . 7

3.3.2 Dynamic Time Warping . . . 7

3.3.3 Examples of Procrustes- and DTW Analysis . . . 7

3.4 Long Short-Term Memory Neural Networks . . . 9

3.4.1 Performance Metrics . . . 10

4 Experimental Setup 11

4.1 Participant Demographics . . . 11

4.2 Test Protocol . . . 11

4.3 Equipment . . . 12

4.4 Inertial Measurement Units . . . 12

4.5 Motion Captured Trajectories . . . 13

5 Method 15

5.1 Preprocessing of Data . . . 15

5.1.1 Preprocessing of Qualisys Data . . . 15
5.1.2 Preprocessing of IMU Data . . . 15
5.2 Synchronization of Motion Signals . . . 16
5.3 Averaging Over Cycles . . . 19
5.4 Feature Extraction . . . 20
5.4.1 Cycle Features . . . 20
5.4.2 Variability-related Features . . . 20
5.4.3 Three-level Classification of Energy Expenditure . . . 20
5.4.4 Propulsion Patterns . . . 21
5.5 Classification of Propulsion Patterns . . . 21
5.6 Implementation of LSTM Neural Network . . . 21
5.6.1 Preparing the Dataset . . . 21
5.6.2 Partitioning of the Data . . . 22
5.6.3 Network Architecture . . . 22

6 Results **24**
6.1 Relation Between Proposed Features and Physiological Values . . . 24
6.1.1 Principal Component Analysis . . . 24
6.1.2 Cycle Features . . . 27
6.1.3 Intrastage Variability . . . 30
6.1.4 Interstage Variability . . . 34
6.1.5 Propulsion Patterns . . . 36
6.2 More on Propulsion Patterns . . . 37
6.2.1 Effect of Speed and Incline on Propulsion Patterns . . . 37
6.2.2 Classification of Propulsion Patterns . . . 38
6.3 Validity of IMU for Cycle Feature Extraction . . . 40
6.4 Reconstruction of Motion Trajectories Using a LSTM Neural Network . . . 43
6.4.1 Performance Metrics . . . 43
6.4.2 Replicating Interstage Variability . . . 44

7 Discussion **47**
7.1 Relation Between Proposed Features and Physiological Values . . . 47
7.1.1 Cycle Features . . . 47
7.1.2 Intrastage Variability . . . 47
7.1.3 Interstage Variability . . . 47
7.1.4 Propulsion Patterns . . . 48

7.2 Validity of Features from IMU Data 48
7.3 Methodological Considerations 49
7.4 Future Work 50

8 Conclusion **51**

References **52**

Appendix **54**

A Plots 54
A.1 Procrustes and DTW Examples 54
A.2 Cycle Features 60
A.3 Intrastage Variability 65
A.4 Interstage Variability 67
A.5 Neural Network-Generated Trajectories 68

List of Figures

2.1 Four classes of propulsion patterns. 3
3.1 Comparison of propulsion patterns using Procrustes analysis. 8
3.2 Comparison of propulsion patterns using DTW analysis. 9
4.1 Test protocol. 12
4.2 Body markers used in the Qualisys motion capture system. 13
4.3 Coordinate frames for the body-mounted IMUs and the Qualisys system. 14
5.1 Qualisys trajectories before and after centering. 15
5.2 Filtering of IMU signals. 16
5.3 Example of good synchronization of IMU and Qualisys data. 18
5.4 Example of bad synchronization of IMU and Qualisys data. 19
5.5 Overview of the neural network architecture. 22
6.1 Biplots of PC1 and PC2 from the control group. 25
6.2 Biplots of PC3 and PC4 from the control group. 25
6.3 Biplots of PC1 and PC2 from the MWU group. 26
6.4 Biplots of PC3 and PC4 from the MWU group. 26
6.5 Cycle features from the wrist marker. 27
6.6 Cycle features from the wrist marker plotted against energy expenditure. 28
6.7 Cycle features from the wrist marker plotted against gross efficiency. 28
6.8 Cycle features from the wrist- and C7 marker plotted against each other. 29

6.9 Intrastage variability metrics from the wrist marker plotted against gross efficiency. 31

6.10 Intrastage variability metrics from the C7 marker plotted against energy expenditure. 33

6.11 Interstage Procrustes distances and differences in gross efficiency. 35

6.12 Boxplots of gross efficiency and energy expenditure grouped by propulsion pattern. 36

6.13 Propulsion patterns for each day-stage combination. 37

6.14 Confusion matrices from pattern classifiers. 39

6.15 Cycle features from IMU and Qualisys data from the control group plotted against each other, before and after synchronization. 40

6.16 Cycle features from IMU and Qualisys data from the MWU group plotted against each other, before and after synchronization. 41

6.17 Motion signals and extracted extremal points, before and after synchronization. . . 42

6.18 Good replication of interstage variability with neural network-generated trajectories. 45

6.19 Bad replication of interstage variability with neural network-generated trajectories. 46

A.1 Comparison of propulsion patterns using Procrustes analysis. 54

A.2 Comparison of propulsion patterns using DTW analysis. 55

A.3 Comparison of propulsion patterns using Procrustes analysis. 56

A.4 Comparison of propulsion patterns using DTW analysis. 57

A.5 Comparison of propulsion patterns using Procrustes analysis. 58

A.6 Comparison of propulsion patterns using DTW analysis. 59

A.7 Cycle features from the wrist marker. Labeled by experiment day. 60

A.8 Cycle features from the wrist marker plotted against gross efficiency and energy expenditure. Labeled by experiment day. 61

A.9 Cycle features from the C7 marker. 62

A.10 Cycle features from the C7 marker plotted against energy expenditure. 63

A.11 Cycle features from the C7 marker plotted against gross efficiency. 64

A.12 Intrastage variability metrics from the wrist marker plotted against EE. 65

A.13 Intrastage variability metrics from the C7 marker plotted against GE. 66

A.14 Interstage DTW distances and differences in gross efficiency from stage 1 to stage 3, labeled by experiment day. Linked points represent a single participant. 67

A.15 Good replication of interstage variability with neural network-generated trajectories. 68

A.16 Good replication of interstage variability with neural network-generated trajectories. 69

A.17 Bad replication of interstage variability with neural network-generated trajectories. 70

A.18 Bad replication of interstage variability with neural network-generated trajectories. 71

List of Tables

4.1 Summary of the characteristics of the participants in the control group. 11

4.2 Summary of the characteristics of the participants in the MWU group. 11

4.3 Experimental information describing the incline and speed (km/h) for each day-stage combination. 12

6.1 Averaged cycle features and physiological values. 30

6.2 Averaged intrastage variability metrics and physiological values. 32

6.3 Averaged interstage variability metrics and difference in gross efficiency between stage 1 and stage 3. 34

6.4 Comparison of averaged cycle features from Qualisys- and IMU data (after synchronization). 43

6.5 Neural network performance metrics. 44

1 Introduction

1.1 Background

Manual wheelchair users (MWUs) face unique challenges and health issues related to their mobility limitations. They are reliant on their upper body in combination with their endurance and propulsion technique to maneuver their wheelchairs, which can result in chronic shoulder pain, joint injuries and other musculoskeletal disorders (Mercer et al., 2006). There is a higher prevalence of physical inactivity in the population of wheelchair users, which leads to increased risk factors for cardiovascular disease and obesity (Wilby, 2019).

Obesity is caused by an energy intake that consistently exceeds one's energy expenditure. To address this challenge, there has been a focus on creating models that predict the energy expenditure in wheelchair users, with the aim of motivating wheelchair users to take active steps towards a healthier lifestyle. In the recent years, low-cost microelectromechanical systems, usually as part of wearable smart-devices, have been taken into use as fitness trackers (Tsang et al., 2016). These methods, however, are often tailored to those without disabilities, and generalize poorly to MWUs. This is due to altered movement patterns and variations in metabolically active muscle mass, making the prediction harder for MWUs (Nightingale et al., 2017).

This thesis investigated frequency, movement patterns and variability in manual wheelchair propulsion and their association with energy expenditure and gross efficiency, with the hope of using these as features in a biofeedback device. This will give the user the opportunity to balance their energy intake to match their energy expenditure, as well the possibility of reducing strain associated with repetitive movements. Through this research, we hope to inspire new interventions and approaches to motivate wheelchair users to lead a healthier lifestyle.

1.2 Research Goals

This thesis sought two primary goals. The first was to explore a set of features based on the frequency of wheelchair propulsion movements and novel features based on the variability in propulsion patterns. The specific objective was to investigate the existence of correlations between these features and physiological values, and the usefulness of these correlations in characterizing the physical exercises performed by wheelchair users. The second goal was to demonstrate how these features can be extracted outside of a lab setting using inertial measurement units (IMUs).

The analyzed features are essentially opportune transformations of the measured movement of wheelchair users during exercise. One of the classes of features we considered are related to quantifying the temporal variability of these movements - in simple words, how repeatable the movements were in time, also depending on the intensity of the exercise. Two different methods for quantifying such variability were then proposed and analyzed. The first variability feature was based on computing the variability in wrist trajectories when the speed and incline of propulsion were constant. The initial hypothesis was that given constant speed and incline, low variability should indicate a more efficient movement pattern, which should be reflected by higher gross efficiency. On the other hand, high variability would suggest that the user is struggling to establish a stable propulsion rhythm, which intuitively should be less efficient. The second variability feature assessed variability with a constant incline but increasing speed. It seems reasonable that the propulsion movement will have to be adapted to the specific speed, so here it was expected to see more variability than in the former case. To explore the relation between this variability and a physiological variable, it was compared to whether there was a difference in the gross efficiency between the lower and the higher speed.

For any of the features analyzed in this thesis to be relevant in a predictive model, we need first to be able to reliably extract them outside a lab setting. We will demonstrate two different methods of how this can be done using equipment that is available in a non-lab setting. The frequency-related features could be directly extracted from acceleration signals, while variability features were based on motion trajectories. To recreate these trajectories, the feasibility of using long

short-term memory (LSTM) neural networks to reconstruct them based on acceleration signals was investigated. The similarity between the reconstructed trajectories and the original ones was then compared.

As secondary goals, this thesis presented tools for classifying propulsion patterns, based on methods from statistical shape analysis and time series analysis. Additionally, we investigated the propulsion patterns and how they change with both speed and incline, with the goal of trying to relate trends to energy expenditure and gross efficiency.

Another goal of the thesis was to investigate how statistically meaningful it is to use data from individuals without disabilities to create predictive models for individuals with disabilities. This investigation follows a practical need - collecting data from people with disabilities tends to be more difficult, since they are a minority of the general population. For this purpose, data from a control group consisting of participants without disability and a group of MWUs was investigated separately, attempting to quantify how similar the observations were for the different groups. The goal was thus to find out whether there were any significant differences between the groups based on the investigated features. If data from the control group generalizes well to the MWU population, it would facilitate the creation of a predictive model by providing a larger data pool.

2 Literature Review

2.1 Propulsion Patterns in Manual Wheelchair Propulsion

Boninger et al., 2002 and Slowik et al., 2015 have amongst others conducted studies examining the propulsion patterns of the arms in manual wheelchair propulsion. Both use the same classification of propulsion patterns into four classes; SC (semicircular), SLOP (single looping over propulsion), DLOP (double looping over propulsion) and ARC (arcing). Figure 2.1 shows propulsion patterns for each of the four classes extracted using motion data from the participants included in this thesis.

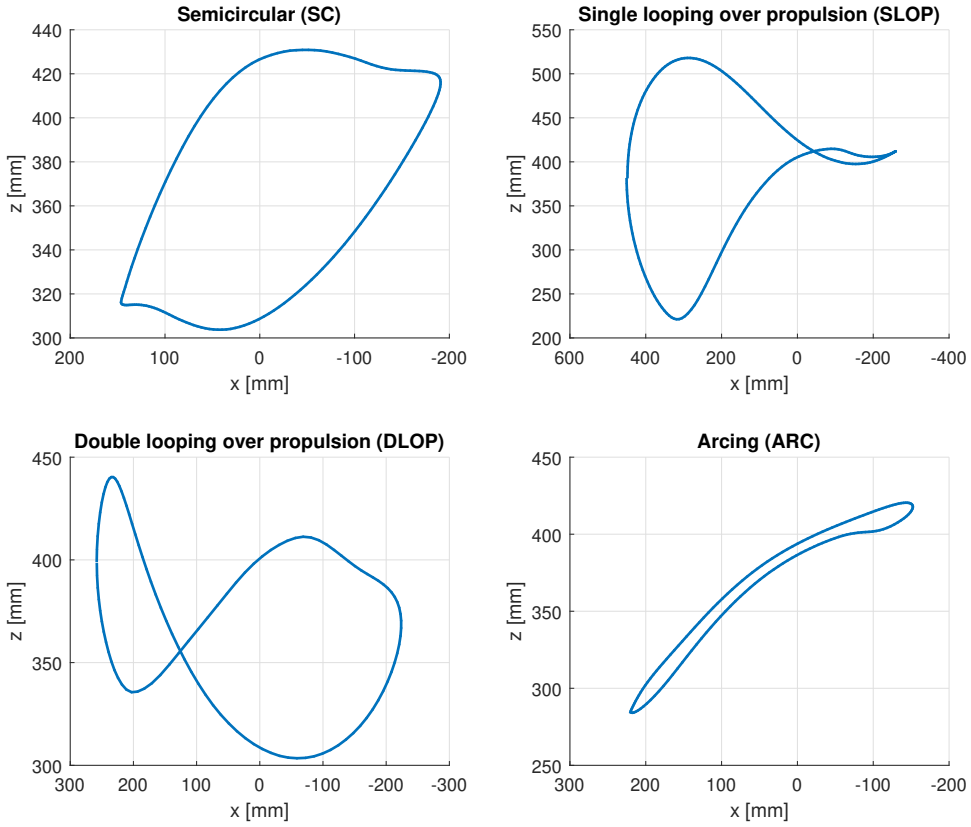


Figure 2.1: Four classes of propulsion patterns.

Boninger et al., 2002 focused on determining whether there were any significant differences in the biomechanics based on the propulsion pattern, comparing various biomechanical variables related to the propulsion with the aim of finding whether there was pattern that reduced the risk of repetitive strain injuries. They found that the SLOP pattern was the most common among 38 experienced wheelchair users. Their conclusion was that the SC pattern was associated with a slower cadence and a longer push phase in the propulsion cycle, proposing that this movement pattern might be more advantageous as there has been shown correlation between the frequency of a task and the risk of injury (Silverstein et al., 1987).

Slowik et al., 2015 investigated the effect of propulsion speed and grade of incline on the propulsion patterns. They collected data from 170 experienced MWUs, and highlight a trend that increasing incline results in a shift towards the ARC pattern. Their analysis suggest that there might not be an optimal propulsion pattern, but rather that the pattern is affected by task-specific constraints and the required upper extremity demand. In another study, Slowik et al., 2016 examined the influence of the propulsion patterns on upper extremity muscle power and stress. Their results suggest that

the DLOP and SC patterns have favourable levels of physicals demand on level ground, and that the use of these can potentially prevent pain and injury during manual wheelchair propulsion.

2.2 Classification of Propulsion Patterns

There have been proposed multiple methods of classifying propulsion patterns into one of the four classes. Koontz et al., 2009 investigates propulsion patterns during start-up propulsion, and utilizes a manual rating system based on the consensus of multiple raters. They discuss that classifying patterns is not as straightforward as it seems, and that distinguishing between patterns is subject to bias of the rater. Some patterns are hybrid of the propulsion classes, and these are especially hard to distinguish. Aissaoui and Desroches, 2008 proposed a classification method based on fuzzy clustering, which reduce the bias of classifying patterns manually. However, the previously mentioned methods rely on having motion captured trajectories for classification, which is complex outside a lab setting. French et al., 2008 proposes a classification method using a wrist mounted accelerometer. They used two machine learning algorithms for classification; k-Nearest Neighbor (kNN) and Support Vector Machines (SVM), both achieving over 90% accuracy.

Whilst not directly related to wheelchair propulsion, Rindal et al., 2018 proposed an automatic classification method of sub-techniques in cross-country skiing using inertial measurement units. They achieved an accuracy of 93.9%, distinguishing between eight classes. They utilized an IMU mounted on the wrist in combination with an IMU mounted on the chest, and used a neural network for classification.

2.3 Variability in Manual Wheelchair Propulsion

Sosnoff et al., 2015 conducted a study examining how the variability in wheelchair propulsion was affected by shoulder pain in the wheelchair user. They conducted several investigations, most concerning the intra-individual variability of the propulsion movement, seeing how kinematic and temporal-spatial parameters varied with time. One specific measure of variability is based on the force profile from a *Smartwheel*, a wheel equipped with force sensors that can be used to detect specific parts of the propulsion cycle. Another measure is based on phase portraits from motion captured trajectories of the arm, quantifying variability by examining fluctuations of the centroid of phase portraits corresponding to individual propulsion cycles. Their findings suggest that individuals with shoulder pain displayed a higher spatial variability than those without shoulder pain in the recovery phase of the propulsion movement.

2.4 Motion Tracking Using Inertial Measurement Units

There has been multiple attempts at both classifying and quantifying human motion using IMUs. If a sufficient number of IMUs are used, it is possible to track the motion of the entire human body (Filippeschi et al., 2017).

Lin et al., 2022 proposed an IMU trajectory tracking system using deep learning to estimate the trajectory of hand- and lower limb movements. They achieved a relative trajectory error (average of the root mean square error in a specified time window) of 0.54 m and highlight the possibility of using IMUs in smart medicine systems, where data from daily activities can be analysed by physicians.

Konak et al., 2020 presented an IMU-based approach for human activity recognition. They used a dataset from six IMUs, which include accelerometer, gyroscope and orientation data. Using a learned statistical model for representing body shapes in natural human poses (Loper et al., 2015) in combination with a pre-trained neural network based on six IMU inputs (Huang et al., 2018), they were able to transform raw IMU signals into 3D trajectories. The trajectories were then flattened to 2D using principal component analysis (PCA) and they were able to generate heatmaps from the trajectories, showing whether some trajectories were more frequent than others. The heatmaps

were classified using a convolutional LSTM neural network. They highlight the influence of drift on the inertial sensors, but mention that this issue can be circumvented by taking advantage of the periodic nature of specific motions such as gait. Their conclusion was that although their method did not achieve state-of-the-art performance on large datasets, their work is an importance example of how transformation of the data to specific domains allow you to take advantage of specific tools that are only available in the specific domains.

3 Theory

3.1 Physiological Values

3.1.1 Energy Expenditure

Values for energy expenditure (EE) used in this thesis was calculated using Weir's formula (Weir, 1949), which is based on internal gas exchange of oxygen $\dot{V}O_2$ $[L/min]$, and carbon dioxide $\dot{V}CO_2$ $[L/min]$:

$$EE = 3.94 \cdot \dot{V}O_2 + 1.106 \cdot \dot{V}CO_2 \quad [kcal/min] \tag{1}$$

3.1.2 Power Output

The power output (PO), also called work rate, is the total work done against the gravitational- and frictional power. Assuming zero air drag, PO can be computed from the body mass m $[kg]$, gravitational acceleration g $[m/s^2]$, rolling resistance coefficient of the wheels μ, incline α, and velocity v $[m/s]$ as follows:

$$PO = m \cdot g(\sin\alpha + \cos\alpha + \mu) \cdot v \quad [W] \tag{2}$$

3.1.3 Gross Efficiency

Gross efficiency (GE) is the percentage ratio between the power output and the total energy expended (Gaesser & Brooks, 1975). It is a measure of mechanical efficiency during exercise, and describes the relation between work output and input. It is computed as follows:

$$GE = \frac{PO}{EE} \cdot 100\% \tag{3}$$

Note that the energy expenditure needs to be in the same unit as the power output, which is transformed as $EE\ [W] = EE\ [kcal/min] \cdot 4.184 \cdot 1000 \cdot \frac{1}{60}[J/kcal \cdot \frac{1}{s/min}]$. Lower energy expended for a given workload means higher gross efficiency, which can be an indication of the individual being better trained or having better techniques.

3.2 Principal Component Analysis

PCA is a dimensionality reduction method commonly used on large datasets containing a high number of features per observation (Jolliffe & Cadima, 2016). The method results in a number of orthogonal principal components (PCs) which are found based on an eigenvalue/eigenvector problem. A simplified view of the method is that it finds the dimension of your dataset which explain most of the variance in the data, then extracts this as the first principal component. After this, the extracted variability is subtracted from the dataset, and a new pass is done in order to find the second principal component, and so on. For each PC we have a loading, which is the coefficients which explains the relative weight of each of the original features in the PC. The linear combination of the original dataset, $\mathbf{X}$, and the loadings, $\mathbf{P}$, results in the scores $\mathbf{T}$ ($\mathbf{T} = \mathbf{XP}$), which can be used to visualize the PCs.

PCA is considered an exploratory data analysis tool and a powerful application for investigating correlation between variables. A specific tool is the *biplot*, which visualizes the scores as points in combination with the loadings as vectors for a set of PCs. This allows us to get an intuition about which of the variables are explained by the specific PCs, the correlation between variables, and whether it is possible to separate specific data points based on specific features.

3.3 Similarity Measures

3.3.1 Procrustes Analysis

Procrustes analysis is a statistical shape analysis method that can be used to find a measure of similarity between two or more shapes (Gower, 1975). The analysis consists of rotation, translation, scaling and reflection of one or multiple shapes to produce the best shape-preserving Euclidean transformation to a reference shape. In our case, we focused on comparing two shapes at a time, which both were motion trajectories corresponding to a cycle of wheelchair propulsion.

3.3.2 Dynamic Time Warping

Dynamic time warping (DTW) is a method used to analyze similarity between time series (Sakoe & Chiba, 1978). It can be used to compare signals of different speeds, and is often used in speech recognition or gait analysis. In gait analysis, motion trajectories from walking at different speeds can be compared to examine whether there's a difference in the gait of the person other than the speed. The method is subject to specific constraints, but the basic idea is to distort the duration of the signals in order produce an optimal match based on a distance metric, commonly Euclidean distance. The method stretches or contracts two time series such that they contain the same number of elements, repeating specific parts of each signal in order to minimize the distance between each corresponding point.

3.3.3 Examples of Procrustes- and DTW Analysis

To get an intuition of how the variability analysis tools work, we have provided a demonstration of them by comparing the archetypal propulsion patterns shown in Figure 2.1. The comparison in this case was done in 2D.

In Figure 3.1 all propulsion classes were compared to the DLOP pattern using Procrustes analysis, and we can see how the method both scales, rotates and moves the shapes in order to find the best match. The Procrustes distances are added in the titles of each subplot. In the lower right we can observe that the method flips the ARC pattern upside down, which from a physical perspective may be an unwanted effect.

In Figure 3.2, all propulsion classes were compared to the DLOP pattern using DTW analysis. To best illustrate the method, we show the individual axes and their value at each step. The signals were normalized to get a common magnitude of the resulting distance metric. The DTW distances are added in the titles here, and we can see how the method stretches out the signals in order to produce the best possible match.

The Procrustes distance metric might be a bit more intuitive as it is in the range $[0, 1]$, while the DTW distance is the Euclidean distance between the warped signals. This means the DTW distance is dependent on the length of the warped signals, and might be harder to get a grasp of. From the figures, we can observe that the two methods give different views of which shapes are most similar. Procrustes finds that DLOP is most dissimilar to SLOP, with a Procrustes distance of 0.12, while DTW finds them to be most similar, with a DTW distance of 133.15. This means that the methods do not capture the exact same variability, which might be helpful in further analysis as including one of the metrics does not lead to the other being redundant.

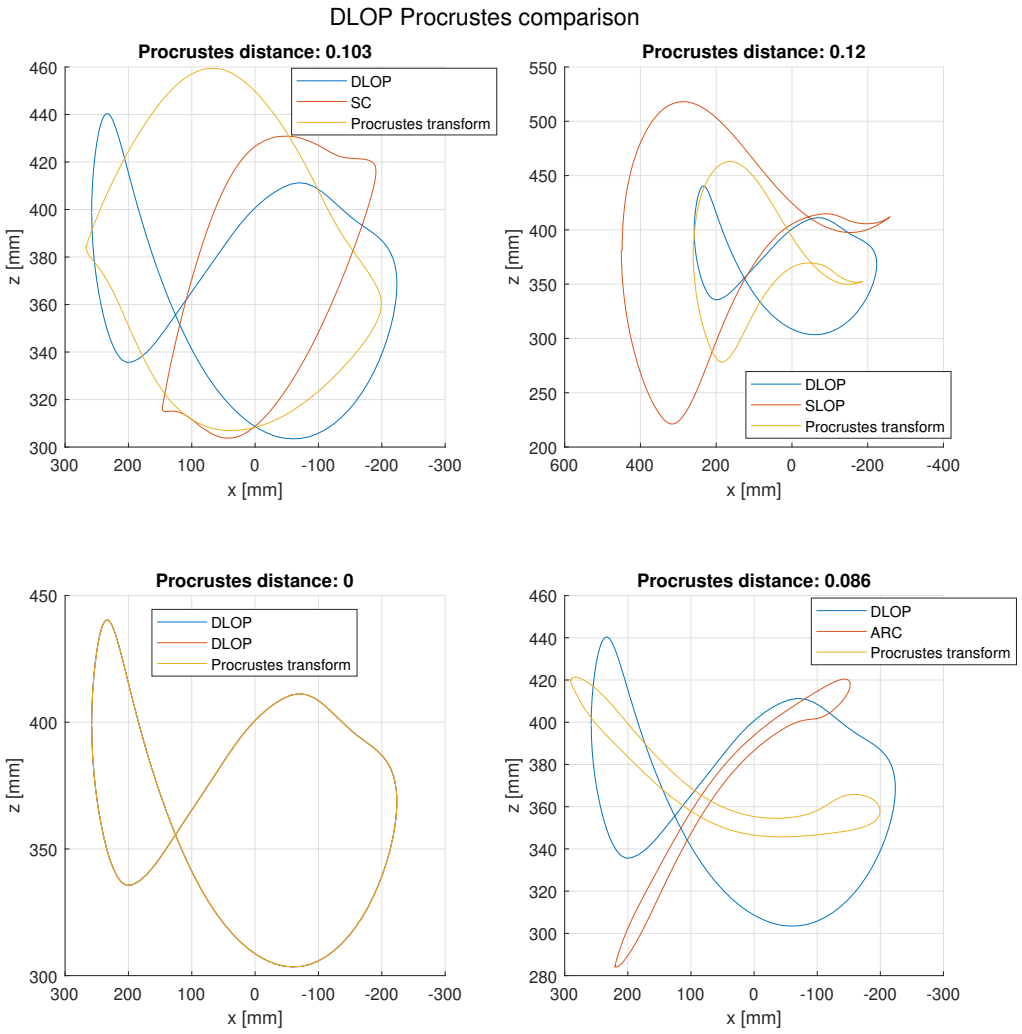


Figure 3.1: Comparison of propulsion patterns using Procrustes analysis. Here, all shapes are compared to the DLOP pattern.

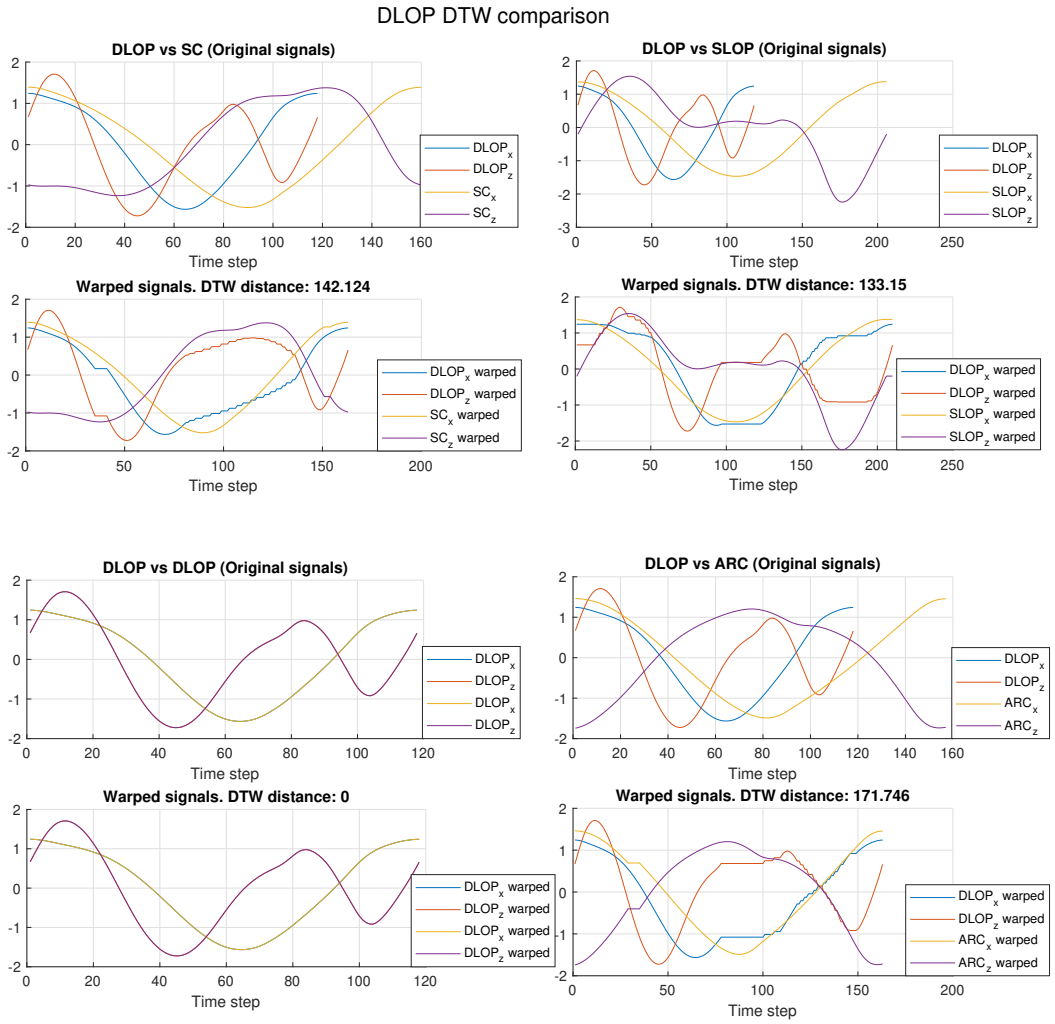


Figure 3.2: Comparison of propulsion patterns using DTW analysis. Here, all shapes are compared to the DLOP pattern.

Plots of Procrustes- and DTW analysis with the other propulsion patterns as a reference is added in Appendix A.1.

3.4 Long Short-Term Memory Neural Networks

A LSTM neural network is a type of recurrent neural networks (RNNs) which is designed to learn long-term dependencies in sequential data (Hochreiter & Schmidhuber, 1997). The key difference between a LSTM neural network and a RNN is that LSTMs incorporate an internal state, which is updated based on inputs and the previous states. More precisely, the LSTM layer of a neural network consists of a *cell*, an *input gate*, a *forget gate*, and an *output gate*. The gates control the flow of information into the cell, which contains information learned from previous time steps. The input gate controls information that may be stored in the internal cell state, and the output gate controls which information from the cell state is passed on by the layer. The forget gate allows the layer to discard information from the previous state.

LSTMs are often used in language processing tasks, speech recognition and in time series prediction. In this thesis, we will use a neural network in an attempt to generate motion trajectories from acceleration signals, exploring various possible architectures.

3.4.1 Performance Metrics

To measure the performance of the neural network, we investigated several metrics. The first metric was the root mean square error (RMSE). Given a vector of observed values $\mathbf{y}$ and a vector of predicted values $\hat{\mathbf{y}}$, the RMSE is defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

The half mean square error (HMSE) was used as the loss function in the neural network. This can be computed based on the RMSE as

$$HMSE = \frac{RMSE^2}{2} \tag{5}$$

Additionally, the Procrustes distance between the estimated trajectory and actual trajectory was used as a metric.

4 Experimental Setup

This section will highlight some of the demographics related to the participants, explain how the data was collected and what equipment was used to collect it. The section is based on Section 3 of the specialization project.

4.1 Participant Demographics

The dataset consists of two parts; data from a control group consisting of 20 participants without disability and data from a MWU group consisting of 15 participants. Data from 20 MWUs has been collected, but only 15 were available at the time of this thesis. Table 4.1 summarizes demographics for the control group, and Table 4.2 summarizes the same demographics for the MWU group.

Gender	Number	Age	Body Mass (kg)	Height (cm)	Body mass index (BMI) (kg/m^2)
Male	9	33 ± 11	81.9 ± 9.4	183.3 ± 5.7	24.4 ± 2.3
Female	11	34 ± 11	67.0 ± 7.8	168.0 ± 5.0	23.7 ± 2.6
Total	20	33 ± 11	75.2 ± 11.4	176.4 ± 9.4	24.1 ± 2.4

Table 4.1: Summary of the characteristics of the participants in the control group.

Gender	Number	Age	Body Mass (kg)	Height (cm)	Body mass index (BMI) (kg/m^2)
Male	10	41 ± 13	81.4 ± 14.7	180.6 ± 9.0	24.8 ± 3.3
Female	5	25 ± 5	70.6 ± 22.9	164.0 ± 5.7	26.3 ± 8.5
Total	15	36 ± 13	77.8 ± 17.8	175.1 ± 11.2	25.3 ± 5.3

Table 4.2: Summary of the characteristics of the participants in the MWU group.

4.2 Test Protocol

Each participant went through three experiment days, which were scheduled within a three-week span. There was a 24 hours minimum period between the experiment days to ensure restitution, and the order of the experiment days for each participant was counterbalanced to reduce biases. Prior to the experiments, demographic variables was collected such as those presented in the tables above. An *International Physical Activity Questionnaire* (IPAQ) (Hagströmer et al., 2006) was conducted to assess the participant's physical activity level, which resulted in one of the categorical scores "low", "moderate" or "high". Additionally, for the MWU group, disability-specific characteristics was collected and whether the participant had a spinal cord injury or not. If the participant had a spinal cord injury, the injury level and ASIA score (Kirshblum et al., 2011) was collected.

An experiment day began with a ten-minute lying resting period and a seated ten-minute resting period where various physiological data was collected. After this, the participant began wheelchair propulsion on a motorized treadmill. The incline of the treadmill was based on the experiment day, and the relation can be seen in Table 4.3. The propulsion phase began with a five-minute warm-up period where the participants were able to familiarize themselves with the setup. After the warm-up was complete, three stages of propulsion were conducted with the speed and inclines shown in Table 4.3. Each stage was four minutes long. Figure 4.1 shows a summary of one experiment day. The figure also shows an incremental test followed by a seated resting period that was only conducted on the 2.5% incline experiment day, from which data will not be included in this thesis.

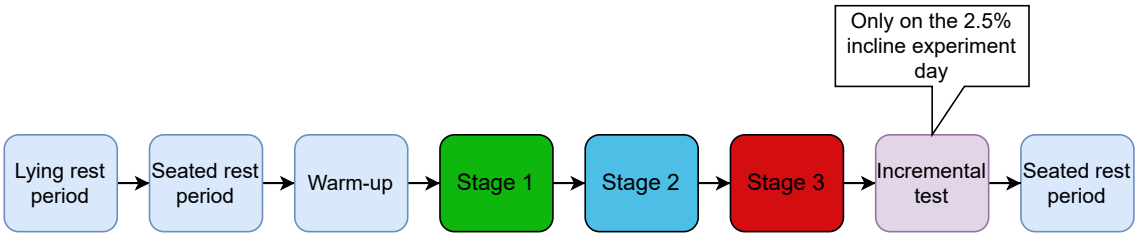


Figure 4.1: Test protocol.

Incline (%)	Day 1 (0.5)		Day 2 (2.5)		Day 3 (5.0)	
	Men	Women	Men	Women	Men	Women
Stage 1	4	3	3	2	2	1
Stage 2	6	5	4	3	3	2
Stage 3	8	7	5	4	4	3

Table 4.3: Experimental information describing the incline and speed (km/h) for each day-stage combination.

4.3 Equipment

The following equipment was used in the experimental setup:

- Motorized treadmill, 6x3m (Bonte Technology, Zwolle, The Netherlands). A safety system was added to the treadmill to constrain the wheelchair movement to forwards and backwards along the treadmill, while preventing the wheelchair from falling off the back of the treadmill.
- Qualisys motion capture system (Qualisys AB, Gothenburg, Sweden). Tracks 3D movement of the markers seen in Figure 4.2, along with four markers mounted on the wheelchair.
- Polar chest strap combined with a Polar M400 HR monitor watch (Polar Electro Oy, Finland). Measures heart rate.
- Five inertial measurement units (IMUs) (Gait Up Physilog®5, Lausanne, Switzerland). Consists of a gyroscope, accelerometer and barometer.
- Vyntus CPX ergospirometer (Vyntus CPX, Vyaire, Medical GmbH, Germany). Measures internal gas exchange, used to compute energy expenditure.

4.4 Inertial Measurement Units

In the experimental setup there were five IMUs; one on the upper back of the participant (next to the C7 vertebrae), one on the chest (on top of the heart rate sensor), one above the left wrist, one under the seat of the wheelchair, and one mounted on the left wheel (7 cm from the center of the wheel). The IMUs mounted on the participants are shown in Figure 4.3a, along with their local coordinate frames. In this thesis we examined data only from the wrist IMU.

From each stage we have four minutes of IMU data, which was captured at 256Hz. The IMUs used in this experiment consist of a triaxial gyroscope measuring in degrees per second, a triaxial accelerometer measuring in g ($1g = 9.81m/s^2$) and a barometer. The IMUs also came with the option of computing euler angles and quaternions which could be used to find the pose of the IMUs. In the previous specialization project there was a focus on gyroscope data from the wrist IMU, but

here we will shift focus to the accelerometer data. This was partly due to the goal of recreating motion trajectories requiring accelerometer data, but also due to the gyroscope data being found to be less reliable. For the stages with low movement intensity, the noise was very prominent in the gyroscope data and made it hard to extract features reliably. The gyroscope data was also found to be prone to drifting, and plotting the orientation of the IMUs using the *poseplot* function in MATLAB showed them to be rotating at various speeds even when lying still.

4.5 Motion Captured Trajectories

The Qualisys motion capture system was used to track the movement of markers that were placed on both the participant and the wheelchair. The movement was measured in millimetres and captured at 120Hz. Figure 4.2 shows all markers placed on the participants. On the wheelchair, there were four markers; one on the center of each wheel and one on a spoke of each wheel. Figure 4.3b shows the three-dimensional coordinate frame for the Qualisys system.

The markers that were used in this thesis are the LFOA- (left forearm) and C7 markers seen in Figure 4.2, as well as the marker placed on the left wheel center. The LFOA marker was the one closest to the wrist IMU, and was used to extract propulsion patterns as well as a reference for trajectories produced from this IMU. Due to the LFOA marker being mounted next to the wrist IMU, this will further be referred to as the *wrist marker*. The C7 marker was mounted next to the upper-back IMU, and was used as an alternate source for the investigated features. The left wheel center marker was used to center data from the rest of the markers as it allowed removal of the variability in trajectories due to movement towards the front or the back of the treadmill.

The Qualisys data was captured twice per stage, roughly at one minute and three minutes into the stage based on a manual button. Each set of Qualisys data consists of 30 seconds of data. Data coming from the first and second window when the motion trajectories were captured per stage will respectively be referred to as the first and second *Qualisys window*.

One issue with comparing motion trajectories to IMU data in the Qualisys windows was the synchronization of data. The Qualisys captures were started based on the manual push of a button, so it is not likely that the captures started at exactly one and three minutes into the stage. A solution for this is presented in Section 5.2.

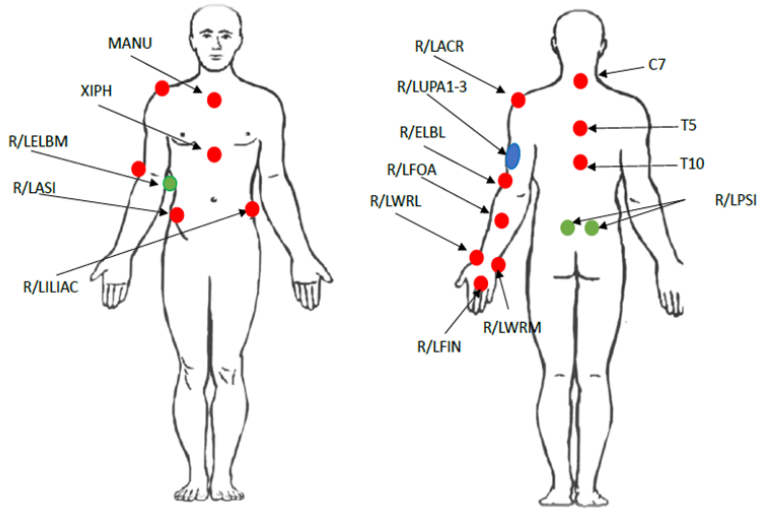
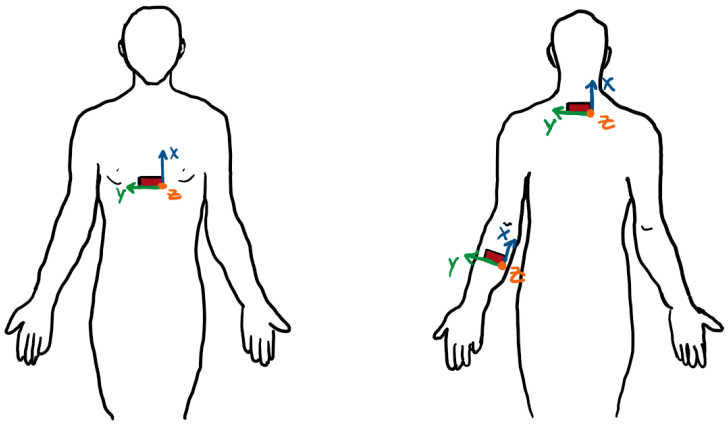
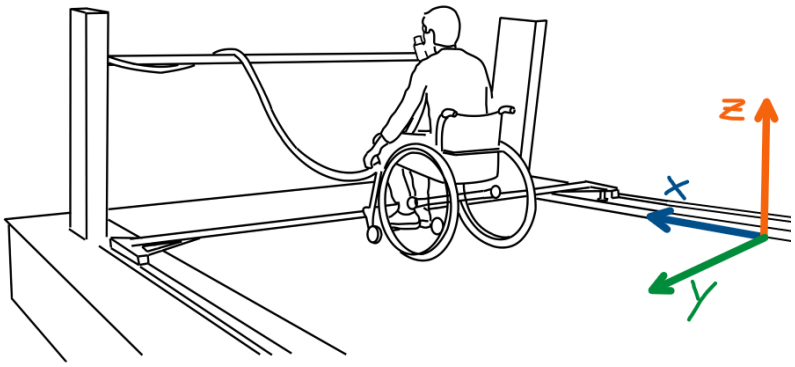


Figure 4.2: Body markers used in the Qualisys motion capture system.



(a) Body-mounted IMUs (Positive Z-axis is out from the figure).



(b) Qualisys coordinate frame.

Figure 4.3: Coordinate frames for the body-mounted IMUs and the Qualisys system.

5 Method

This section will elaborate which preprocessing steps were necessary and how features were extracted from the data.

5.1 Preprocessing of Data

5.1.1 Preprocessing of Qualisys Data

The extraction of some features required different preprocessing of the Qualisys data, and is described in detail in the sections regarding the specific features. In all cases, however, the variation along the treadmill was removed by centering the data to the left wheel center. This was done by subtracting the position of the left wheel center from the position of all markers at all time steps. This allowed us to ignore displacement along the treadmill, and it became easier to compare trajectories between both sessions and participants. An example of motion trajectories before and after centering is shown in Figure 5.1.

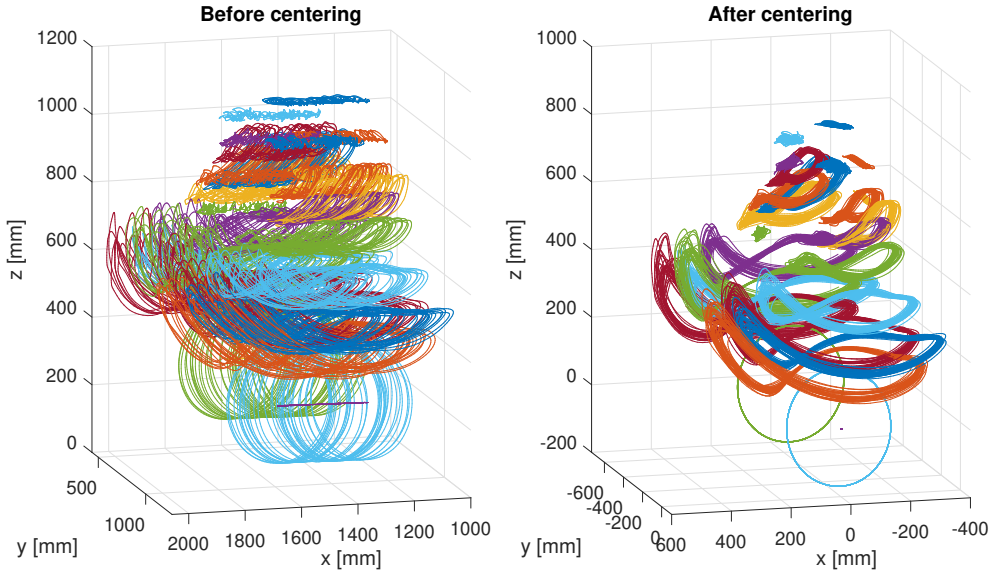


Figure 5.1: Qualisys trajectories before and after centering.

5.1.2 Preprocessing of IMU Data

To facilitate comparison between the IMU- and Qualisys data, the IMU data was downsampled to have the same frequency as the Qualisys data. Upsampling the Qualisys data was another possibility, but this was found to introduce noisy edge effects, and thus downsampling was favoured.

As the IMU data contained a lot of noise, various filtering approaches were tested. The initial approach was to use a Butterworth lowpass filter, which introduced a time lag to the signal. This made comparison between signals harder, which we wanted to be as accurate as possible. Another approach was to use a Gaussian filter. This was inspired by Rindal et al., 2018, which used the same brand of IMUs as we do. The Gaussian filtering removed most of the high-frequency noise without delaying the signals, and was chosen as the favoured filtering method. A comparison of the filtering methods on an acceleration signal is shown in Figure 5.2.

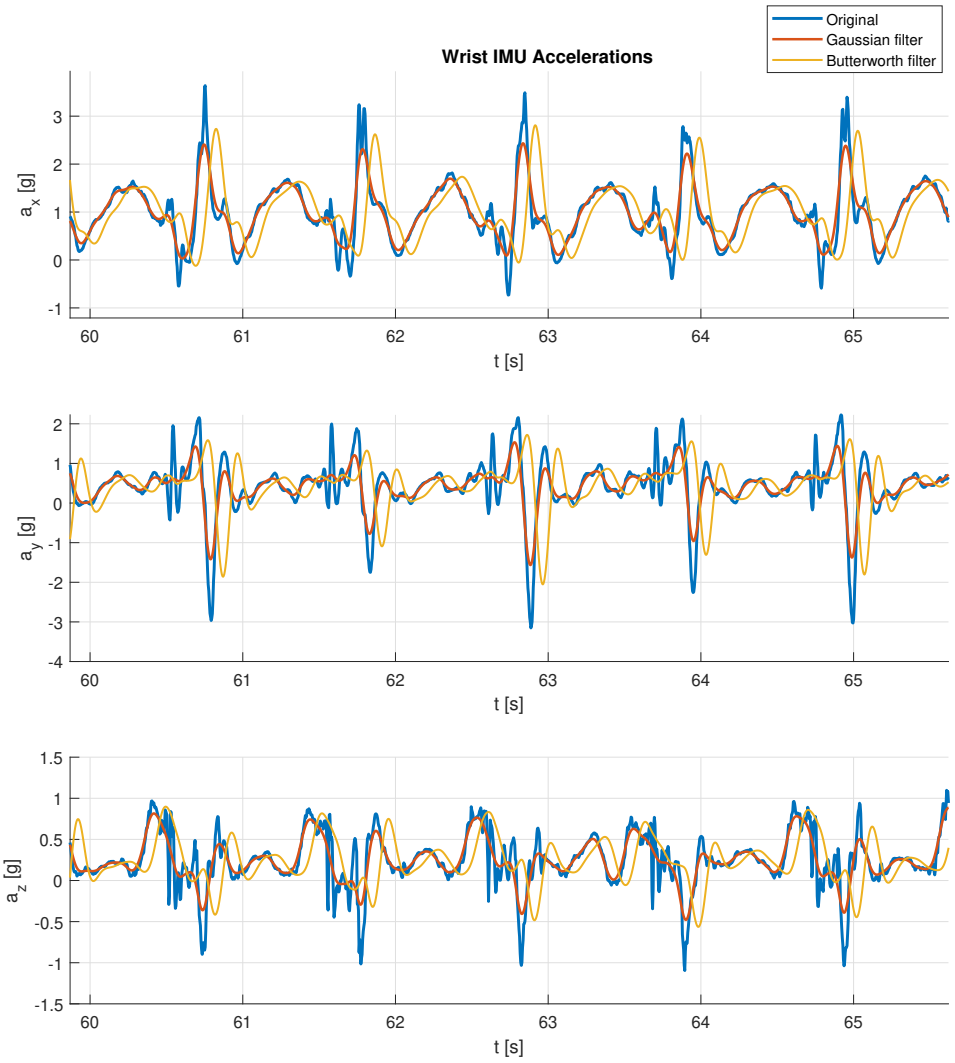


Figure 5.2: Filtering of IMU signals.

5.2 Synchronization of Motion Signals

A big challenge in this thesis was to produce the most accurate comparison between Qualisys- and IMU data. As mentioned in Section 4.5, the Qualisys data was captured twice within a propulsion stage based on the manual push of a button. This made it hard to accurately extract IMU data in the exact windows where we had Qualisys data as a reference. A proposed solution to this was to compare timestamps between the signals in order to find the exact locations of the Qualisys windows in the IMU data. However, the timestamps of the Qualisys data were based on the clock of a lab computer, which after the data collection was found to be out of sync with the clock related to the IMUs. An idea was to find out whether there were a constant bias term between the clocks, which was not the case. Comparison of the clocks over the course of multiple weeks showed that the bias was drifting, and thus comparison of timestamps was not a feasible solution.

The approach that made synchronization of the signals possible was based on exploiting the alignment of the wrist IMU with the Qualisys coordinate frame. Under manual propulsion, it was observed that the X-axis of the IMU mounted on the wrist was approximately aligned with the Z-axis of the Qualisys coordinate frame. An ad hoc solution was to extract acceleration from the Qualisys data by double-differentiating it, then applying a moving average filter to reduce noise

from the differentiation. After this, we were able to compare the signals in order to find the exact overlap. The method was implemented using the *alignsignals* function in MATLAB, which estimates the delay between two signals based on the cross-correlation of the signals. Both signals were normalized in order for them to be on the same scale, and to remove the gravity component from the IMU acceleration. To make things easier, the 240s IMU acceleration was split into two parts, where we aimed to find the first and second Qualisys window in the respective parts. This was possible because the first window is generally between 60-90s and the second window between 180-210s. Due to the repetitive nature of the propulsion, this allowed us to reduce the possibility of a mismatch. The result given a correct match is that we knew the exact starting point of each Qualisys window relative to the start of the IMU data.

To automate the synchronization process as much as possible, we utilized some inclusion and exclusion criteria. Based on the timestamps from the Qualisys data, we could find the difference in seconds between the start of the first and second Qualisys window. If the difference was within a 0.5 second threshold of the difference found using the *alignsignals* method, the signals were deemed synchronizeable. Otherwise, the signals needed to be manually controlled. If the delay between the Qualisys data and the IMU data was zero or negative, the signals were deemed unsynchronizeable using the current method.

An example of good signal synchronization is shown in Figure 5.3. The upper part of the figure shows the entire IMU signal with identified Qualisys windows, and the two lower plots shows respectively the first and second Qualisys window. The method found that the first Qualisys window started at 60.6 seconds and the second at 180.6 seconds. These are within a second of the expected starting points of each Qualisys window, and additionally we can observe that the difference is equal to the difference from the Qualisys timestamps.

On the other side, Figure 5.4 shows an example of bad signal synchronization. From the two lower plots we can see that the acceleration extracted from the Qualisys data has little resemblance to the original IMU acceleration. The method identifies the Qualisys windows to start at respectively 90.8 seconds and 175.2 seconds, which does not seem reasonable according to the experimental setup. The cause of the signals being so dissimilar is unknown, but a possible reason may be a mounting error of the IMU which would lead us to looking at the wrong IMU-axis. Another possibility is that the assumption that the IMU X-axis and Qualisys Z-axis aligns does not hold for specific propulsion methods.

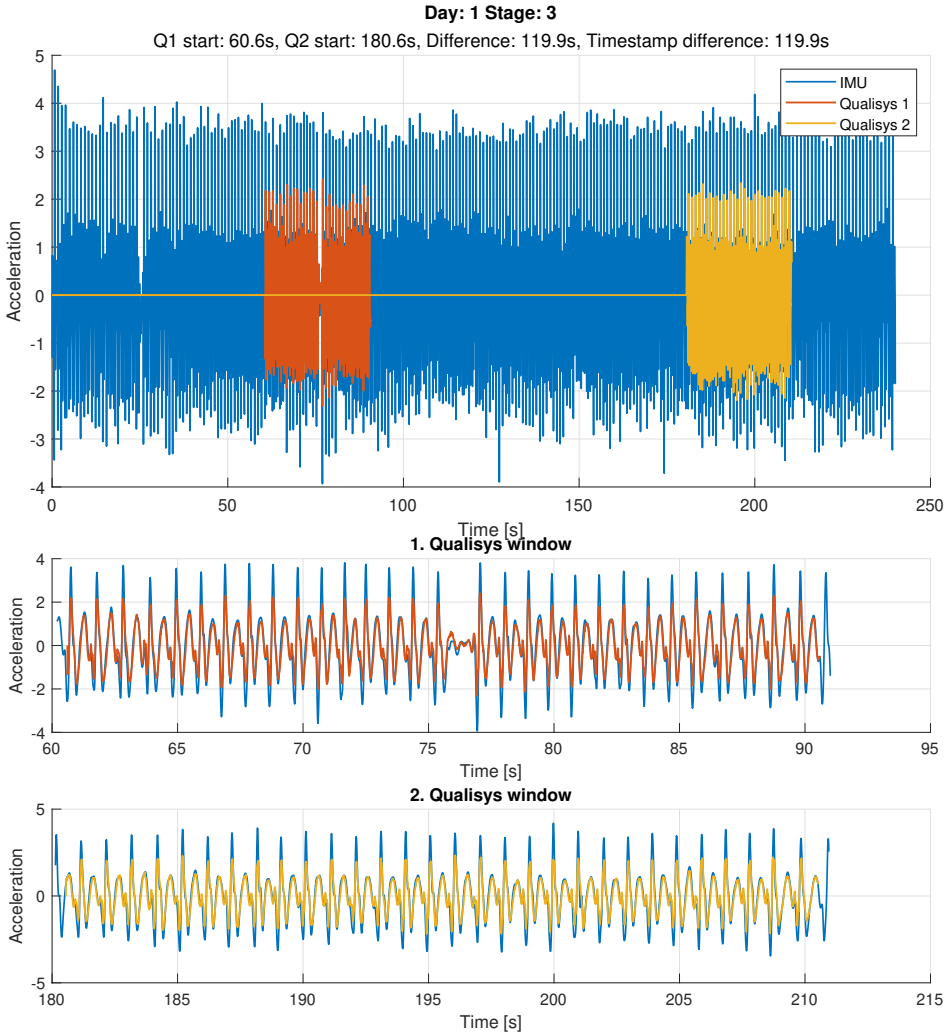


Figure 5.3: Example of good synchronization of IMU and Qualisys data.

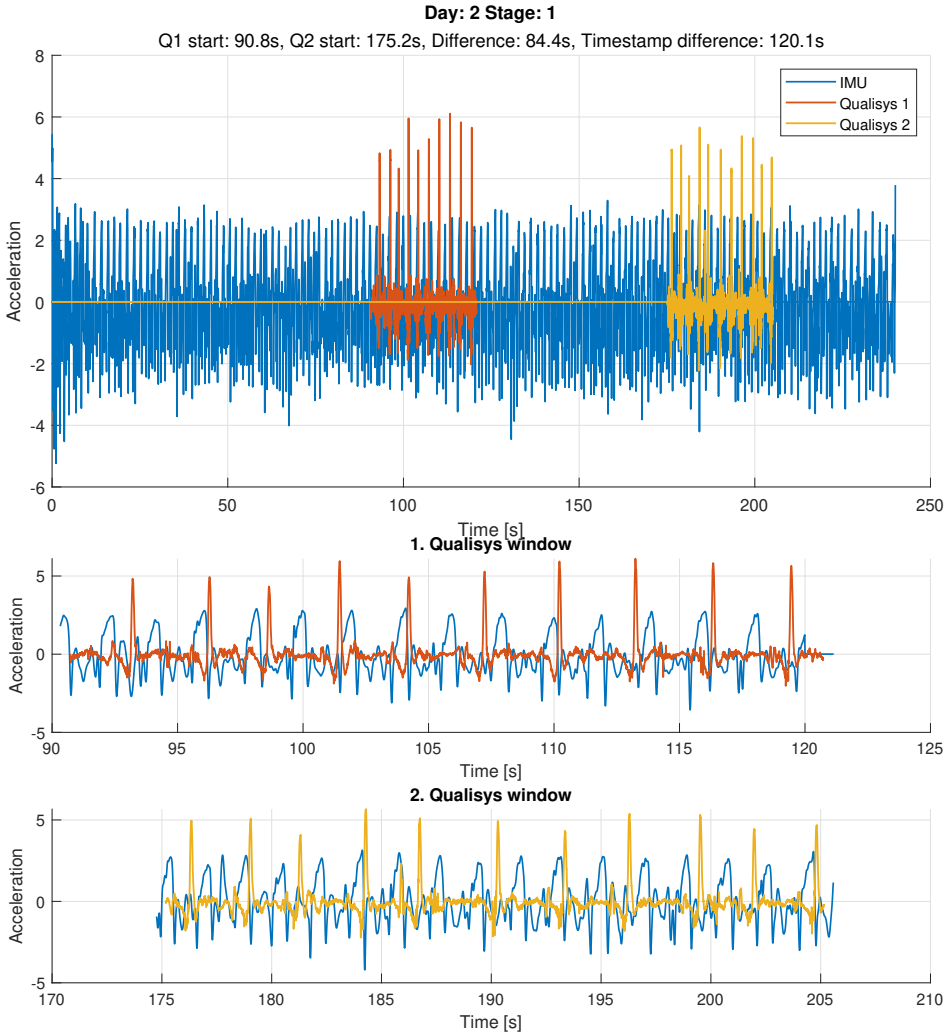


Figure 5.4: Example of bad synchronization of IMU and Qualisys data.

5.3 Averaging Over Cycles

Given a trajectory that contains multiple cycles, we were able to create an average cycle that captured the characteristics of the propulsion pattern. This method was based on extracting extremal points from the trajectories, which is further described in Section 5.4.1, and using these to split the trajectories into individual cycles. In order to be able to compare the shape of averaged trajectories using either Procrustes- or DTW analysis, we needed the averages to have approximately the same start and end points. To achieve this, we used the maximas in the X-direction of the Qualisys frame for the splitting. Use of other extremal points was also possible, but these maximas tends to be close to where the X-maximas of the wrist IMU are, so they are used for easier comparison. An example of extracting extremal points is later shown in Figure 6.17.

Given the individual cycles, we needed to decide on a length of the average cycle as the length of each individual cycle may wary. For this, the median cycle length was used. This was chosen in order to make the method more robust to misclassified cycles that may pull an average length up or down. After finding the median cycle length, each individual cycle was interpolated to have this length, and finally we took an average over all cycles at each time step.

Although the designed method is robust to misclassified cycles in terms of the cycle length, a misclassified cycle will still have a negative effect on the generated average. For instance, if a classified cycle contains two or more cycles, taking the average at each step can lead to the misclassified cycle giving points at the wrong end of the cycle. This effect has a larger significance for sessions with fewer number of cycles, as taking the average with many cycles reduces the effect of misclassifications. Due to this, it was important to detect cycles as accurately as possible.

5.4 Feature Extraction

5.4.1 Cycle Features

The extraction of cycle features is described in more detail in Section 4 of the specialization project, but will be summarized shortly here.

Extraction of the cycle features, i. e. the cycle times and number of cycles within a Qualisys window, were based on using the *findpeaks* method in MATLAB. This was the case for both Qualisys and IMU data, the difference being the threshold specified in the method. For Qualisys data, we detected maxima along the X-axis after subtracting the mean, using a threshold of 40 (mm) and a minimum time of 0.7s between maxima. For IMU accelerations, we used the same minimum distance and a threshold of 0.4 (g) to find maxima along the X-axis, also subtracting the mean here. If there were detected less than five cycles in the acceleration signals, the threshold was lowered to 0.2 (g). The reason for this adaptive threshold is due to the method failing to detect cycles for very low intensity sessions.

5.4.2 Variability-related Features

In order to investigate the variability of propulsion patterns, Procrustes- and DTW distances were used as metrics. Procrustes distances were computed using the *procrustes* function in MATLAB, which also returned the transform between the compared patterns. DTW distances were computed using the *dtw* function, which returns the warping path between signals, allowing extraction of the distorted signals.

The variability was examined in two cases; *intrastage* and *interstage*. In both cases, averaged propulsion cycles were compared.

The intrastage variability was computed by comparing trajectories between the first- and second Qualisys window within a stage, seeing how the averaged propulsion cycle changes over time given constant speed and incline.

The interstage variability was computed by comparing trajectories between stages. For simplicity, only trajectories from the second Qualisys window of a stage was compared, between the first and the last stage of an experiment day. This variability gives a metric to quantify how the trajectories change with increasing speed and thus intensity. Computing the interstage variability was only possible given that there was data available from both of these stages for the given experiment day.

5.4.3 Three-level Classification of Energy Expenditure

In the previous specialization project we used the metabolic equivalent of task (Jetté et al., 1990) as a metric for the energy expenditure of a person relative to their mass, which could be divided into five intensity classes. This is a metric commonly used for people with no disabilities, but is not sufficiently valid for wheelchair users due to disuse or paralysis in the lower limbs resulting in reduced muscle mass (Nightingale et al., 2017). There has been attempts to adapt the MET metric to wheelchair users with spinal cord injuries (Collins et al., 2010), but we have instead used another approach.

In order to investigate the relation between our features and energy expenditure, we utilized a three-level classification of intensity based on the energy expenditure. This was done by first computing the steady state energy expenditure for all stages by taking the average of the energy expenditure during the final 60 seconds of each stage, ten seconds offset from the end of the stage to reduce transient effects. Having computed the steady stage energy expenditure, we sorted all values for all participants and divided them into the three equally big classes *Low*, *Moderate* and *High*.

5.4.4 Propulsion Patterns

Using the propulsion classes described in Boninger et al., 2002 and Figure 2.1 as a reference, averaged propulsion patterns from each set of Qualisys data was manually labeled into one of the four classes. As mentioned in Section 2.1, this manual labeling method is subject to bias of the rater and there might be cases where the propulsion patterns are hybrid of the classes. In these cases the pattern was labeled into the class best characterizing the pattern.

5.5 Classification of Propulsion Patterns

Using the archetypical patterns for the given classes as pictured in Figure 2.1, averaged trajectories were compared to these using tools from Procrustes- and DTW analysis, then classified as the class which gave the smallest distance.

5.6 Implementation of LSTM Neural Network

The goal with using a neural network was to be able to recreate the motion trajectories outside of a lab setting. Initially we wanted to test a similar approach to Konak et al., 2020 for generating the trajectories, but this required gyroscope and orientation data as well as the acceleration. Due to the sensor- and orientation drift mentioned in Section 4.4, this approach was considered unviable. Instead, we will examine an approach based solely on acceleration data.

The neural network was implemented using the MATLAB Deep Learning Toolbox and its LSTM implementation (MathWorks, n.d.-a), and was trained on a GPU (NVIDIA GeForce RTX 3070).

In the following subsections we will explore different network architectures and propose different partitions of the data in order to test the generalizability of the network.

5.6.1 Preparing the Dataset

Since this was a supervised learning approach, we needed to use the IMU data in the Qualisys windows in order to have a reference trajectory. In this case, both the Qualisys- and IMU data were normalized. The Qualisys data was measured in millimetres, and the specific ranges of values for trajectories depended on the height and arm-length of the participant. Normalizing the trajectories allowed us to compensate for this effect, and it also helped to convert the IMU and Qualisys signals to a common scale. Normalizing the signals also allowed the neural network to process inputs more effectively, reducing computational time (Sola & Sevilla, 1997).

If we were able to use all data from all participants, we would get 35 $(participants) \cdot 3$ $(days) \cdot 3$ $(stages) \cdot 2$ $(Qualisys\ windows) = 630$ sets of data, each consisting of data from a single Qualisys window. However, 80 sets of data were excluded due to the synchronization issues described in Section 5.2. In order to get the most out of the data that we were able to include, we proposed a windowing method where each set of data was split into multiple windows consisting of a specified amount of propulsion cycles. The required cycles for a window was a tuneable parameter. After some trial and error it was decided to use seven cycles as a minimum value, which increased the available data by a factor of ≈ 2.5. The same method as described in Section 5.4.1 was used to

detect the cycles, then the data was split into windows containing the specified amount of cycles. If there were not enough cycles to fill an entire window, these cycles were added to the previous window. This was done in order to ensure that each window was robust to misclassification of cycles.

The data that was extracted in each window was the acceleration, the reference trajectory, and an averaged trajectory that was generated from the reference trajectory in the specific window. All data was normalized.

5.6.2 Partitioning of the Data

We investigated three different methods of partitioning the dataset. The first two methods involved excluding data from respectively whole experiment days and whole stages. In both cases, the excluded data was used as test sets in order to see whether removing some days or stages created a bigger impact on the model performance. The non-excluded data was divided into training and validation sets with 80% of the data for training and 20% for validation. The reason for using most of the non-excluded data for training was due to the test set containing approximately a third of all data, so we wanted to expose our model to as much data as possible while still keeping some for validation. The third method was to split the control and MWU groups. Each group was used in entirety as the training set in this case, and the other group was split into validation and test sets. Since there were more participants in the control group than the MWU group at the time of this thesis, ten participants were used in the test set in both cases, and the rest were used as the validation set. The participants were split randomly into the validation and test sets, and we used two random splits in each group for cross-validation.

5.6.3 Network Architecture

The final implementation of the neural network is visualized in Figure 5.5. The input is 3D accelerations, and the output is 3D motion trajectories.

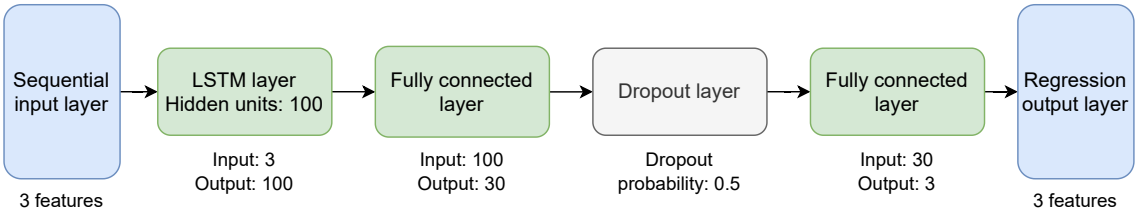


Figure 5.5: Overview of the neural network architecture.

The sequential input layer, regression output layer as well as the last fully connected layer were essential for the network architecture as they specified the input- and output dimensions, but the rest of the layers could be both modified and removed. We will now look at the effect of changing these layers.

Increasing the number hidden units in the LSTM layer to 200 resulted in a higher training-validation gap, with approximately the same performance on the training set. Reducing the number to 50 led to worse RMSEs, but approximately the same training-validation gap as with 100. Removing the LSTM layer altogether resulted in much worse performance, and all output trajectories tended to be a flat circle. This isn't surprising as it is the LSTM layer that contains an internal state which allowed the network to learn dependencies between time steps.

Increasing the output size of the first fully connected layer led to a small increase in performance on the training set, but not with the same increase on the validation set. Decreasing the output size made the training-validation gap smaller, but reduced the ability of the network to catch the complexity of the trajectories. Removing the layer altogether didn't have a big impact on the

performance, possibly due to the other fully connected layer being able to compensate for the loss in complexity.

Removing the dropout layer resulted in better performance on the training sets, but led to a high gap in performance between training and validation sets. The dropout layer randomly sets input elements to zero during training with the specified dropout probability, and will prevent the network from memorizing the training data. The gap in performance showed that removing the dropout layer made the network overfit the training data, which led to worse performance on both validation and testing sets.

In addition to the specified layers, a number of training options for the network could be added. The ones added are summarized in the following list:

- Solver: Adam (adaptive moment estimation). A gradient descent optimization algorithm with adaptive learning rates.
- Max epochs: 50. An epoch is the full pass of the training algorithm over the training set.
- Learning rate: Initially 0.001, but is multiplied by 0.8 every 10th epoch. The reason behind using this is to optimize the training process by allowing the model to make bigger adjustments in the beginning of training and smaller towards the end.
- Gradient threshold: 2. By limiting the magnitude of the gradient when training, we aim to improve the stability and speed of the learning process. Exploding gradients can lead to slow convergence and can have a large impact on the weights on the neural network during training.
- L2 regularization: 0.0001. Adds a penalty term to the loss function of the model based on the L_2 norm, with the goal of improving model robustness and preventing overfitting of the training data.
- Validation patience: 4. If the computed loss on the validation set does not improve for the specified number of times, training of the network is stopped early. This is included to prevent the network from overfitting the training data.

6 Results

6.1 Relation Between Proposed Features and Physiological Values

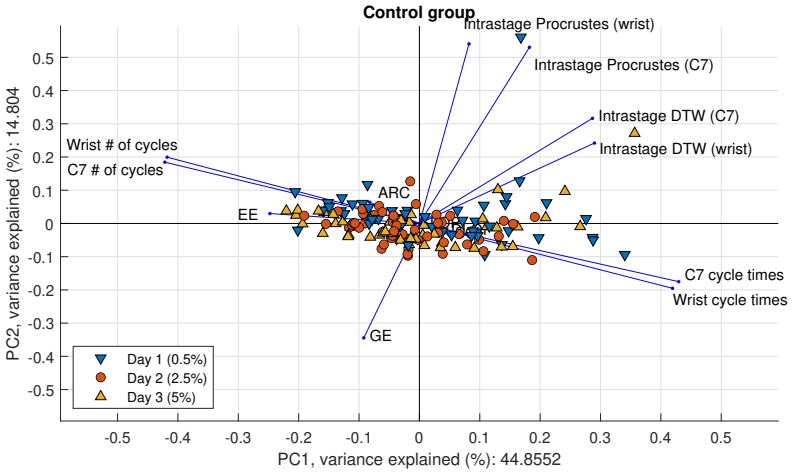
In the following subsections we will first conduct PCA in order to visualize all features and their possible relation to physiological values, then take a deeper dive into each feature.

6.1.1 Principal Component Analysis

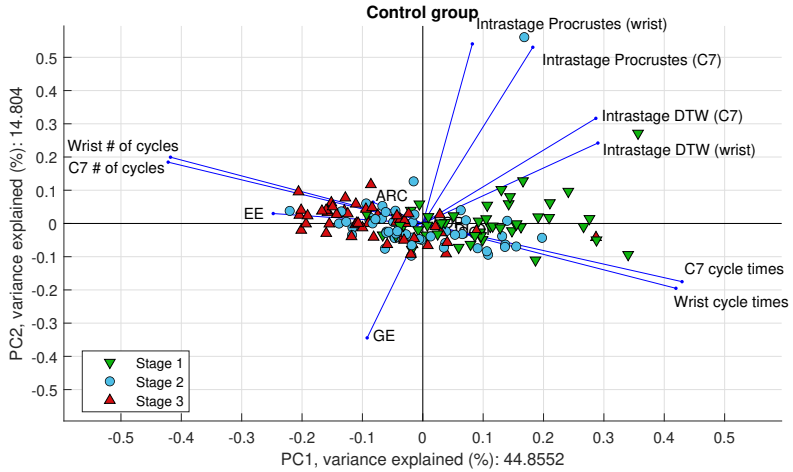
To get an initial overview over the potential relevance the proposed features in predicting energy expenditure or gross efficiency, PCA was conducted on a dataset consisting of all proposed features as well as energy expenditure and gross efficiency with the exception of the interstage variability feature. The interstage feature was not included as it compares data between stages, which would require us to modify or exclude all other features to have a sensible comparison. The control- and MWU group was analysed separately in order to see if the same correlations could be observed in each group, or whether there were any differences. Since data with different units and magnitudes were compared, the dataset was normalized. The propulsion patterns were added using one-hot encoding, i. e. by adding a separate column to the dataset for each pattern, then setting a "1" for the relevant pattern for the current observation and "0" for the others. The columns containing one-hot encodings were not normalized.

Starting with the control group, we can observe in Figure 6.1 that PC1 explains almost half of the variance in the control group data. The features that contribute the most here are the cycle features as well as energy expenditure. The number of cycles are strongly anticorrelated to the cycle times, which makes sense as in order to fit in more cycles in a 30 second window you need faster cycles. The energy expenditure seems to be correlated with the number of cycles, which seems rational as more movement means more energy is required. PC2 explains about 15% of the variance, with the intrastage variability features being dominant. There seems to be correlation between all of the variability features, but the correlation between the Procrustes distances and the DTW distances separately seems strongest. Intuitively it would perhaps make more sense that the wrist Procrustes- and DTW distances are more strongly correlated with each other than the same feature from the C7 marker. As we saw in Section 3.3.3, the Procrustes- and DTW distances resulted in different views of which shapes were the most similar, and this is a possible explanation. The gross efficiency seems to be anticorrelated to the variability features, which can imply that less variability in your movement means a higher efficiency. For PC3 and PC4, we can see in Figure 6.2 that there seems to be stronger correlation between the wrist variability metrics, and the same for C7. The C7 variability metrics seem to be anticorrelated to the energy expenditure, which might indicate that sessions that have more variability in the upper body is correlated to lower energy expenditure. PC3 and PC4 explain a much smaller portion of the variance compared to PC1, however, and the correlations found here might not be as significant.

In Figures 6.1a and 6.2a the points are labeled by the experiment day, and there seems to be no obvious clustering of points in either of the biplots. This isn't that strange as each experiment day ranges from low to high intensity, so the points will represent the entire ranges for all features. By labeling points by the stage however, Figure 6.1b shows signs of clusters along the PC1-axis. Stage 1 has higher cycle times than stages 2 and 3, and stage 2 is centered between the others. This implies that the cycle features are of relevance in determining the energy expenditure due to the correlation observed between them. This is supported by the energy expenditure generally being higher for the more intense stages, which seems to be possible to separate based on the cycle features. There are also some signs of clustering in the biplot of PC3 and PC4 in Figure 6.2b, but with more overlapping between the groups.

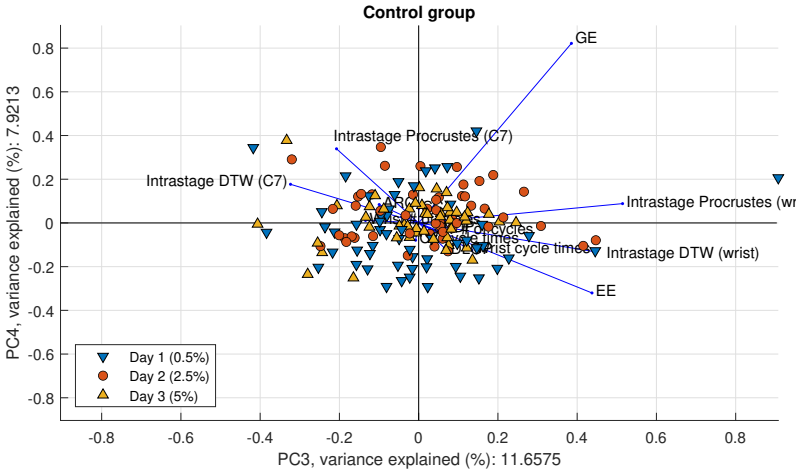


(a) Labeled by day.

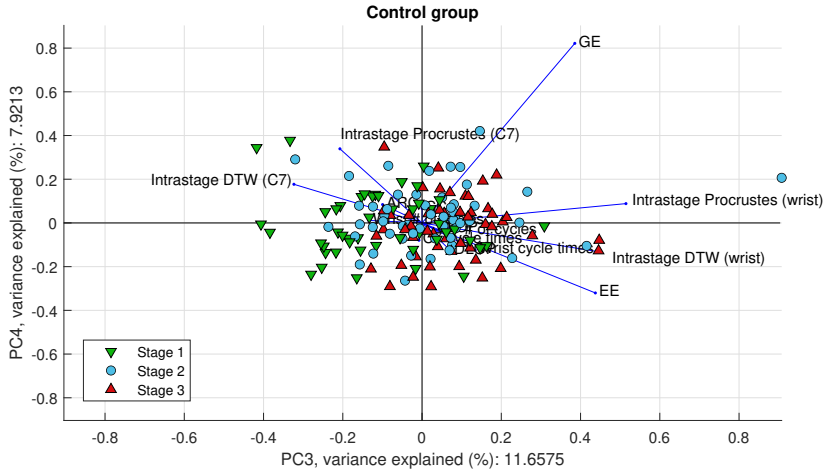


(b) Labeled by stage.

Figure 6.1: Biplots of PC1 and PC2 from the control group.

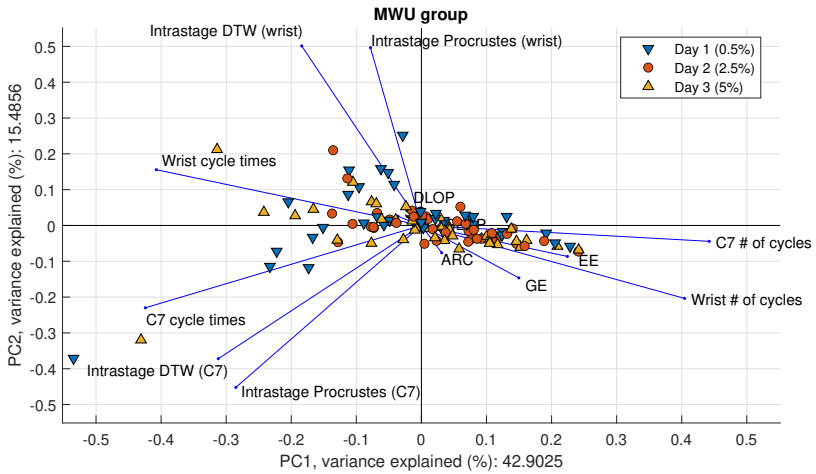


(a) Labeled by day.

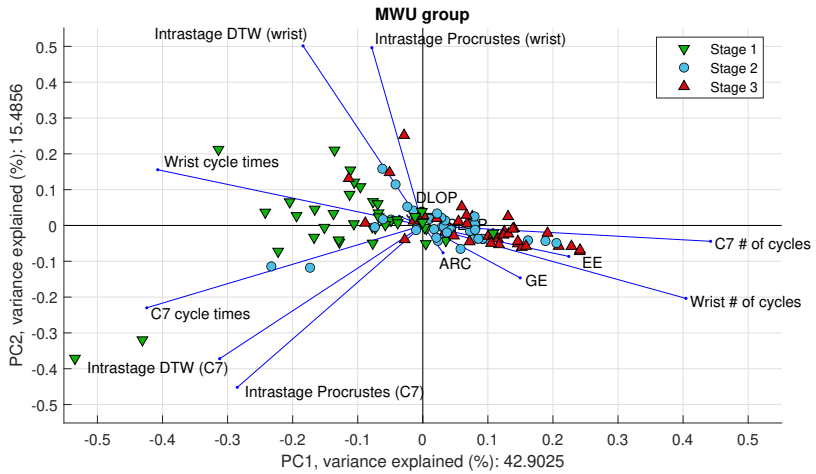


(b) Labeled by stage.

Figure 6.2: Biplots of PC3 and PC4 from the control group.

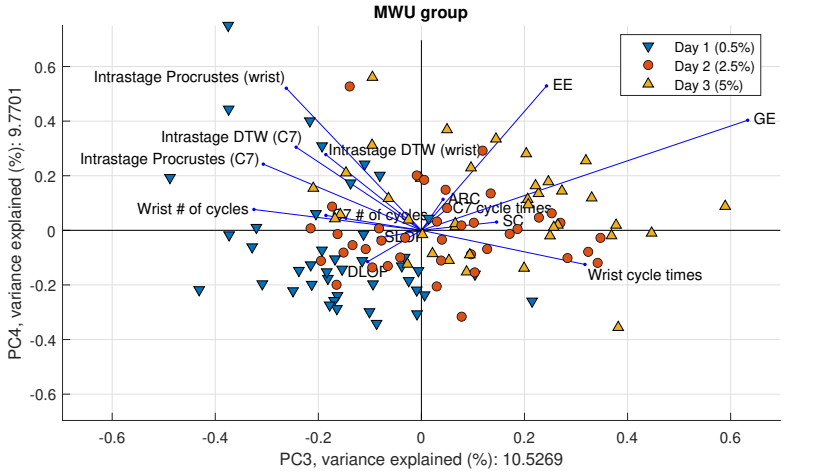


(a) Labeled by day.

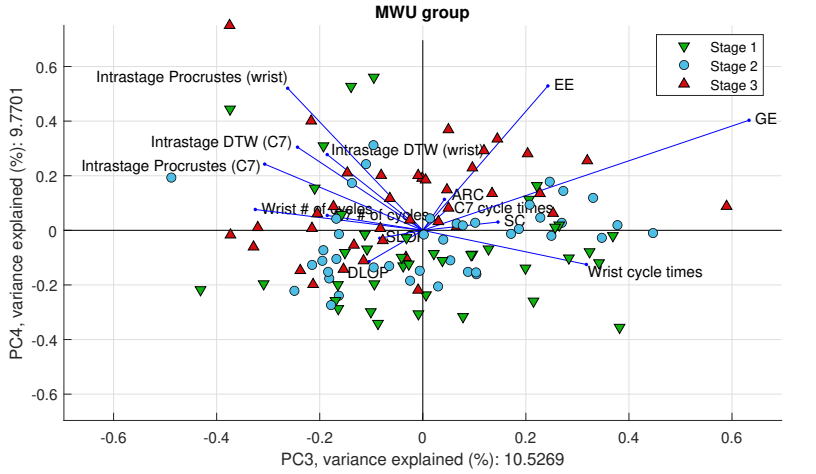


(b) Labeled by stage.

Figure 6.3: Biplots of PC1 and PC2 from the MWU group.



(a) Labeled by day.



(b) Labeled by stage.

Figure 6.4: Biplots of PC3 and PC4 from the MWU group.

For the MWU group, we can see in Figure 6.3 that the cycle features and energy expenditure also are important in PC1, though the loading vectors for cycle features from the C7- and wrist marker are not as close together as in the control group. However, they have approximately the same contribution in PC1, which can be seen by projecting the vectors onto the PC1-axis. In PC2, the variability features are prominent as well, though here we have more signs of correlation between the variability metrics from each source separately. There is not the same correlation between the variability metrics from the wrist- and C7 markers as we saw for the control group. Regarding PCs 3 and 4, we can see in Figure 6.4 that the loading vectors of the variability features from each of the markers are more close to each other, and can also see signs of correlation between the energy expenditure and gross efficiency.

When labeling by experiment day, Figure 6.3a shows little clustering in PC1 and PC2. However, Figure 6.4a shows signs of clustering. The clusters seems to be along the dimension where the energy expenditure and gross efficiency is most significant, with day 3 having highest values. This makes sense given that day 3 with 5% incline is more physically demanding than the others, leading to higher energy expended. However, this helps little in the prediction as there is little correlation between the features and the physiological values in these PCs. Points towards the upper left are correlated with high intrastage variability, and we can see that points from day 1 are mostly located here.

Upon a closer look of the center of Figure 6.4a, it seems that the DLOP pattern is more prominent in day 1, and the ARC and SC patterns are more prominent in the other days. It is worth mentioning though that these pattern do not contribute as much as other features in either of the PCs, which is reflected by the short loading vectors.

By labeling by the stages, Figure 6.3b shows clustering similar to the control group in the biplot of PC1 and PC2. For PC3 and PC4, Figure 6.4b shows less obvious clustering.

For both the control- and MWU group, the propulsion patterns contribute little to any of the first four PCs, which may imply that the other features are more important in prediction of physiological variables. Another reason might be that the PCA method is best suited for continuous variables, causing the one-hot encoding of the categorical variables to have little impact on the variance in the dataset.

6.1.2 Cycle Features

Figure 6.5 shows the averaged cycle features from the wrist marker for the control- and MWU group. Here we can see that the cycle features are indeed anticorrelated with a slight nonlinear trend. While the stages are not fully separable, there is a trend that higher cycle times and lower number of cycles are more prominent in the lower-speed stages, and vice versa for the high-speed stages.

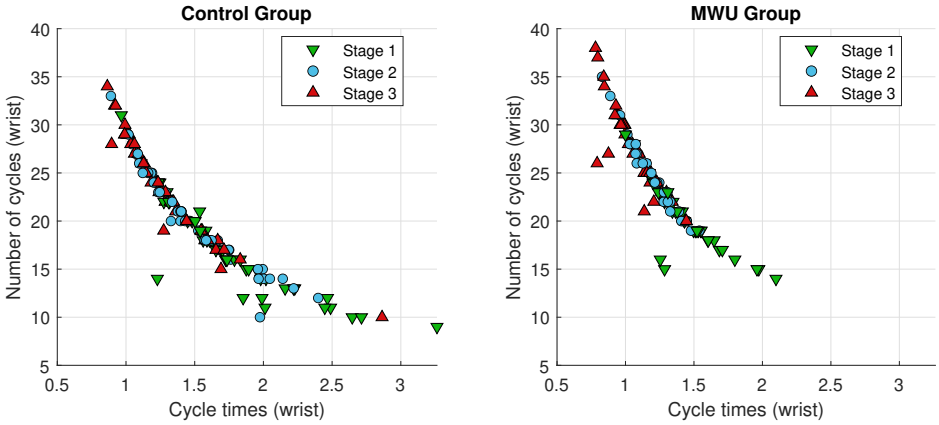


Figure 6.5: Cycle features from the wrist marker. Labeled by stage.

Adding another dimension to the plots above, Figure 6.6 shows the cycle features plotted against energy expenditure. Points from stage 1 seems to be easier to separate than for stages 2 and 3, which are overlapping more. There is an obvious increase in energy expenditure as the cycle times get lower and number of cycles get higher, which implies that the cycle features are relevant in predicting energy expenditure.

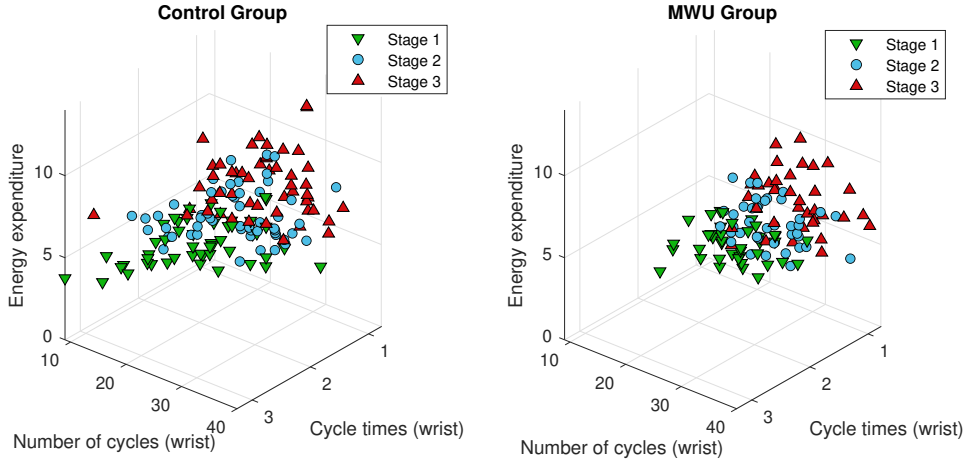


Figure 6.6: Cycle features the from wrist marker plotted against energy expenditure. Labeled by stage.

Similarly, Figure 6.7 shows the cycle features plotted against gross efficiency. Its harder to separate the stages here than when comparing the cycle features to energy expenditure. However, there is still a notable difference between stage 1 and the other stages, especially in the control group.

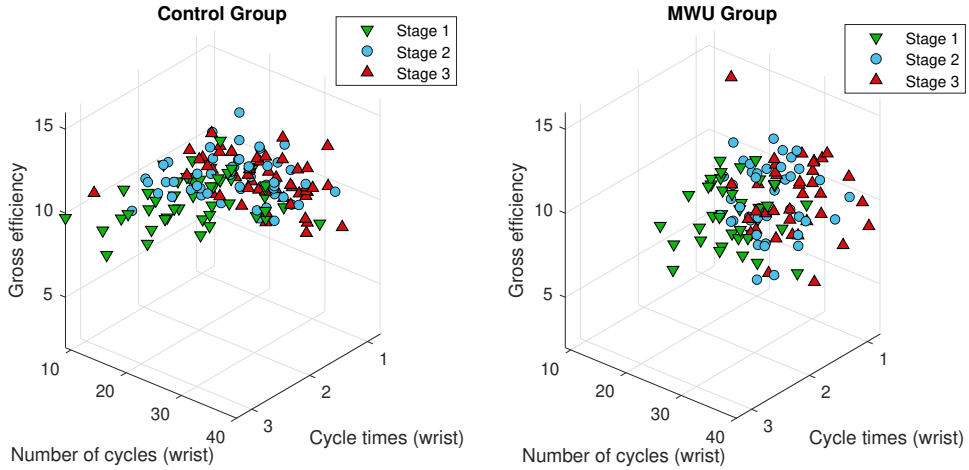


Figure 6.7: Cycle features from the wrist marker plotted against gross efficiency. Labeled by stage.

In this section only cycle features from the wrist marker labeled by stage was shown. Plots of cycle features labeled by experiment day as well as corresponding plots with cycle features from the C7 marker are included in Appendix A.2. The correlation between cycle features from the wrist- and C7 marker is illustrated in Figure 6.8. For moderate and high intensity, the cycle features are approximately the same between markers. For low intensity, however, the cycle times from the C7 marker are often higher.

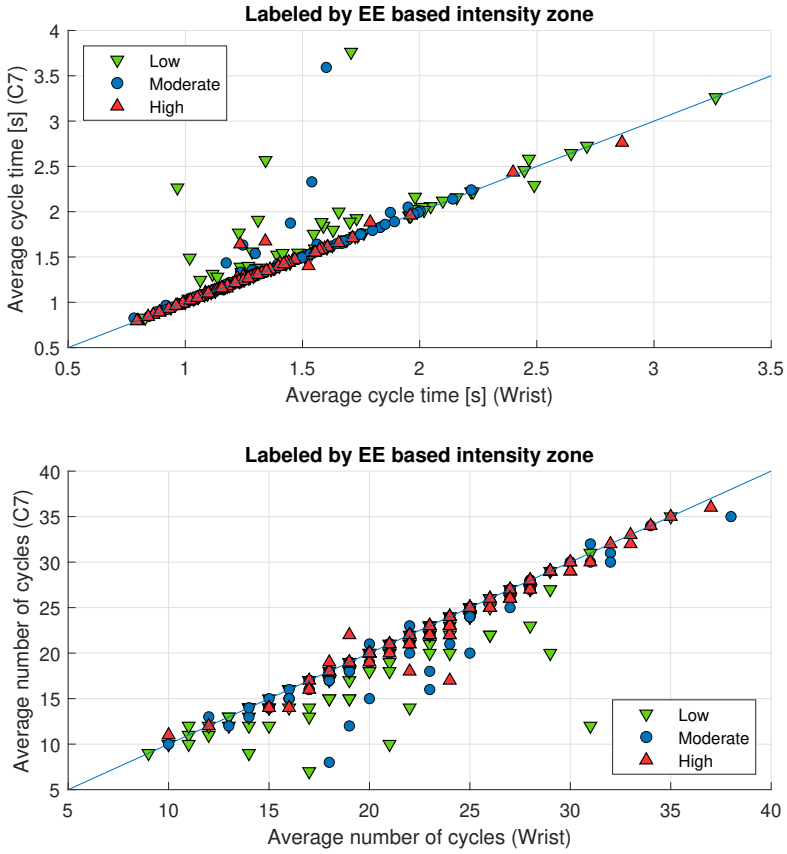


Figure 6.8: Cycle features from the wrist- and C7 marker plotted against each other. Labeled by energy expenditure-based intensity.

Table 6.1 shows the average values of the cycle features from both markers for each day-stage combination. For each group, there is a trend that both the cycle times and number of cycles converge to the same value for stage 3 of all days, though the value differs between the groups. The MWU group generally has lower cycle times, and converge towards a cycle time which is approximately 0.2 seconds faster than the control group. Features from the C7 marker are often offset from features from the wrist in stage 1, though they converge to the same values in the higher stages. The cause of this may be that there is a lower physical demand in the lower stages, which is reflected in less movement of the upper body. Another possible cause could be that the cycle detection method is not as accurate for the low-intensity stages for the C7 marker.

		Cycle Times (s)		Number of Cycles		EE (kcal/min)	GE (%)
		Wrist	C7	Wrist	C7		
Day 1	Stage 1	1.82	2.02	16.50	13.89	3.30	8.70
	Stage 2	1.46	1.53	21.00	19.68	4.86	9.68
	Stage 3	1.28	1.27	24.72	24.50	7.05	8.90
Day 2	Stage 1	1.61	1.68	18.81	17.44	4.14	9.60
	Stage 2	1.46	1.50	20.56	19.78	5.49	10.42
	Stage 3	1.33	1.33	22.33	21.83	6.98	10.58
Day 3	Stage 1	1.80	1.85	16.63	15.38	4.50	8.71
	Stage 2	1.47	1.49	20.65	19.94	6.36	10.05
	Stage 3	1.25	1.25	23.87	23.60	8.26	10.52

(a) Control group.

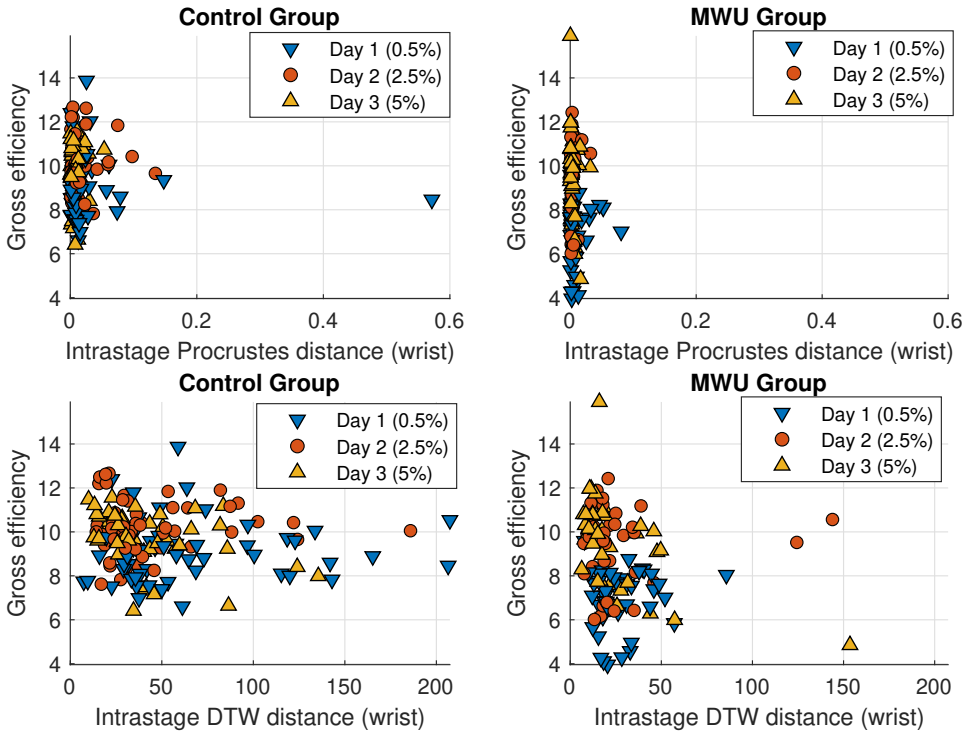
		Cycle Times (s)		Number of Cycles		EE (kcal/min)	GE (%)
		Wrist	C7	Wrist	C7		
Day 1	Stage 1	1.39	1.71	20.86	17.64	2.94	6.49
	Stage 2	1.21	1.27	24.60	23.47	4.14	7.25
	Stage 3	1.06	1.09	27.31	26.62	5.62	7.16
Day 2	Stage 1	1.35	1.37	21.92	21.31	3.84	8.87
	Stage 2	1.19	1.19	24.79	24.43	4.89	9.50
	Stage 3	1.08	1.08	27.31	27.23	6.17	9.48
Day 3	Stage 1	1.56	1.72	18.62	17.23	4.44	8.51
	Stage 2	1.24	1.24	23.92	23.62	6.19	9.73
	Stage 3	1.07	1.07	27.25	27.00	7.84	10.32

(b) MWU group.

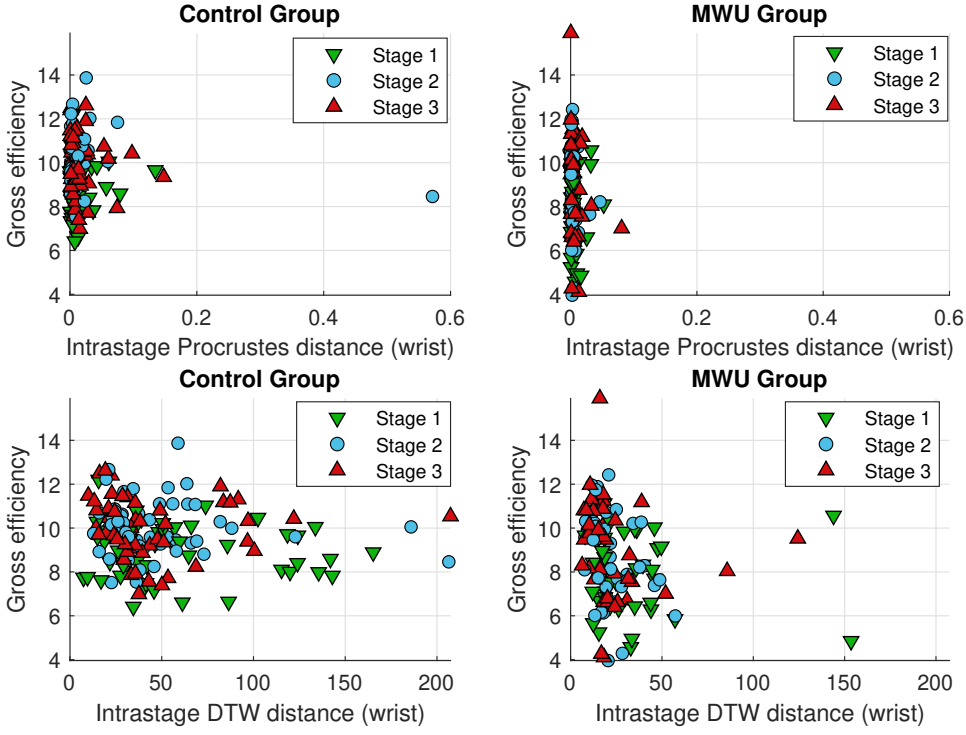
Table 6.1: Averaged cycle features and physiological values.

6.1.3 Intrastage Variability

From the PCA biplots in Figures 6.1 and 6.3, there seemed to be a slight anticorrelation between the intrastage variability features and the gross efficiency, especially for Procrustes distances in the control group. Table 6.2 sums up the intrastage metrics for both groups, and Figure 6.9 shows the intrastage features from the wrist marker plotted against gross efficiency. There seems to be no obvious patterns in either the plots labeled by experiment day or the plots labeled by stage, though there is an indication of a trend that the points with highest gross efficiency have low intrastage variability. The variability is generally lower for the higher incline days and higher speed stages, though the difference is only a few decimals in the Procrustes case. The trend is more apparent in the DTW case, where points are generally more spread out than in the plots with Procrustes distances. The cause of this may be due to the DTW distances being affected by the length of the sequence. Since the cycle times from stage 1 are generally higher than for the other stages, cycles extracted from this stage will contain more points which may in turn result in a larger cumulative distance between compared signals.



(a) Labeled by experiment day.



(b) Labeled by stage.

Figure 6.9: Intrastage variability metrics from the wrist marker plotted against gross efficiency.

		Procrustes distance		DTW distance		EE (kcal/min)	GE (%)
		Wrist	C7	Wrist	C7		
Day 1	Stage 1	0.021	0.062	77.19	213.15	3.30	8.70
	Stage 2	0.039	0.048	56.89	139.83	4.86	9.68
	Stage 3	0.023	0.023	58.30	79.87	7.05	8.90
Day 2	Stage 1	0.018	0.029	40.06	139.88	4.14	9.60
	Stage 2	0.015	0.029	47.67	104.18	5.49	10.42
	Stage 3	0.018	0.009	45.01	98.55	6.98	10.58
Day 3	Stage 1	0.009	0.109	55.69	200.06	4.50	8.71
	Stage 2	0.009	0.022	37.08	82.42	6.36	10.05
	Stage 3	0.009	0.013	30.22	85.71	8.26	10.52

(a) Control group.

		Procrustes distance		DTW distance		EE (kcal/min)	GE (%)
		Wrist	C7	Wrist	C7		
Day 1	Stage 1	0.009	0.101	28.32	127.78	2.94	6.49
	Stage 2	0.009	0.056	27.45	83.92	4.14	7.25
	Stage 3	0.013	0.020	28.50	59.29	5.62	7.16
Day 2	Stage 1	0.005	0.025	34.44	113.40	3.84	8.87
	Stage 2	0.003	0.008	17.45	70.53	4.89	9.50
	Stage 3	0.006	0.008	27.11	72.74	6.17	9.48
Day 3	Stage 1	0.007	0.065	38.32	155.03	4.44	8.51
	Stage 2	0.003	0.008	20.32	70.08	6.19	9.73
	Stage 3	0.004	0.012	14.91	71.80	7.84	10.32

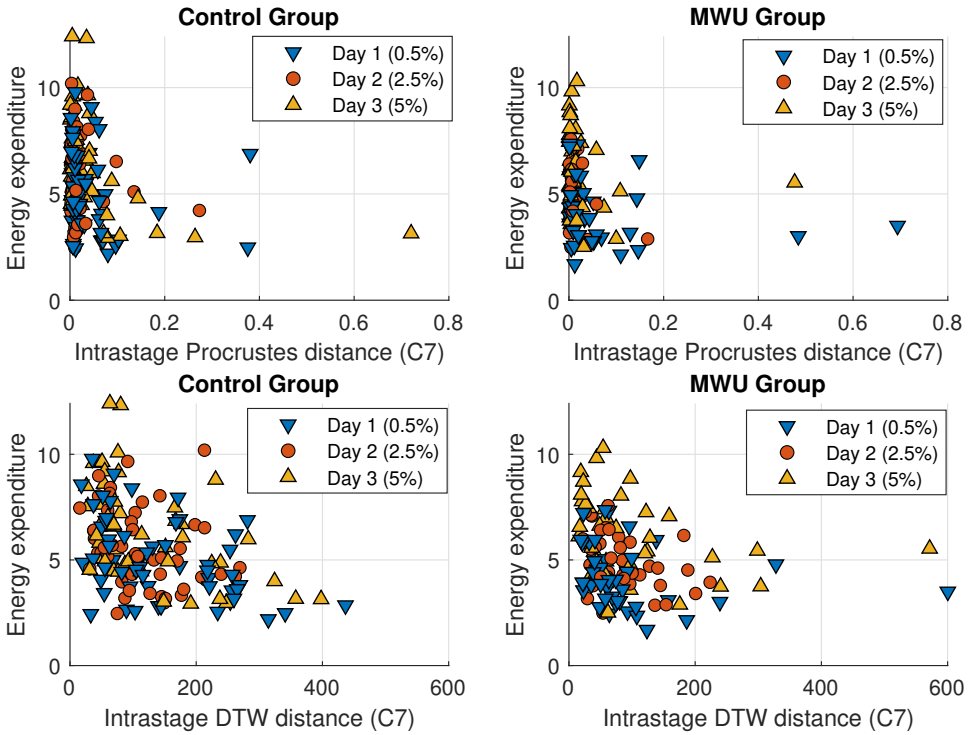
(b) MWU group.

Table 6.2: Averaged intrastage variability metrics and physiological values.

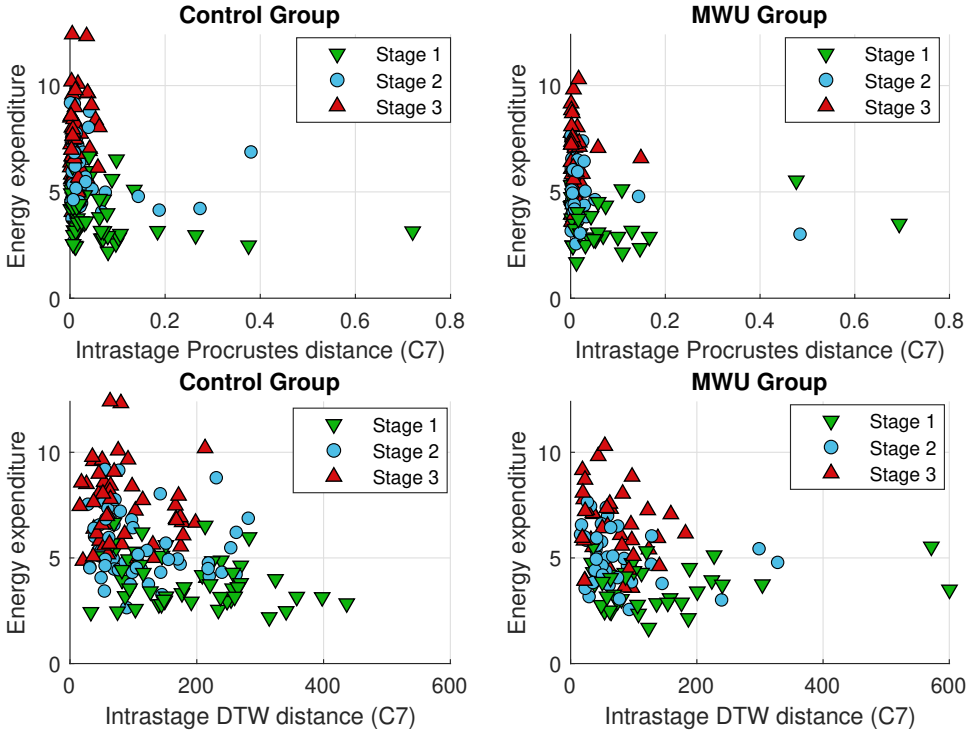
In Figure 6.2 we saw an anticorrelation between the intrastage C7 metrics and the energy expenditure. Figure 6.10 shows the intrastage variability for the C7 marker plotted against energy expenditure, and there are signs of clustering in the stages. Most of the clustering, however, is along the energy expenditure dimension, which is obvious given that each stage is more demanding than the previous one. However, there seems to be a trend in the DTW distances that the distances are larger for stage 1 and smallest for stage 3 in both the control- and MWU group. While it is hard to do any exact separation based on the intrastage metric, it may be useful together with other features in predicting the energy expenditure.

In Table 6.2, there is a notable difference in variability between the wrist- and C7 markers in the first stage, independent of the incline. This is possibly due to the low intensity of the first stage allowing the wheelchair user to move their upper body in a less constrained manner. For the higher stages, we can see that the variability decreases, and in the third stage the variability is approximately equal to the wrist variability. This is likely due to the increased speed requiring the wheelchair user to have a more stable movement pattern.

Plots of intrastage variability from the wrist marker plotted against energy expenditure and intrastage variability from the C7 marker plotted against gross efficiency are added in Appendix A.3.



(a) Labeled by experiment day.



(b) Labeled by stage.

Figure 6.10: Intrastage variability metrics from the C7 marker plotted against energy expenditure.

6.1.4 Interstage Variability

Figure 6.11 illustrates the interstage variability metric plotted against the *difference* in gross efficiency using Procrustes distances as the variability metric. The difference is computed by subtracting the gross efficiency from stage 3 from the gross efficiency from stage 1, meaning that a positive difference represents an increase in gross efficiency from stage 1 to stage 3. Points that are linked represent a single participant. We can see a trend that day 2 generally has the smallest interstage variability, and that day 1 and 3 has a noticeably higher variability for most of the participants. This is further confirmed by Table 6.3, which shows the interstage variability metrics summed up along with the differences in gross efficiency. The effect can be noticed in the variability from both the wrist- and C7 marker, although the C7 variability has bigger differences.

There is a notable difference in how much the gross efficiency increases as the incline gets higher, with points from days 2 and 3 generally having a higher increase than day 1. This is more clear in the control group, where there is generally little difference in the gross efficiency in day 1. However, it is hard to do any separation of the experiment days based on the interstage variability due to most of the separability being in the difference in gross efficiency, which is what we wanted to predict. If we assume that we know the incline, this feature might be useful in determining the efficiency of the wheelchair user as there was a more notable increase in gross efficiency for steeper inclines.

A plot of the interstage variability using DTW distance as a metric is added in Appendix A.4.

	Procrustes distance		DTW distance		GE difference
	Wrist	C7	Wrist	C7	
Day 1	0.086	0.108	134.97	286.06	0.137
Day 2	0.058	0.052	71.54	159.96	1.133
Day 3	0.062	0.146	84.72	203.35	2.118

(a) Control group.

	Procrustes distance		DTW distance		GE difference
	Wrist	C7	Wrist	C7	
Day 1	0.056	0.120	80.46	176.40	0.493
Day 2	0.048	0.078	52.84	134.48	1.070
Day 3	0.080	0.134	69.15	204.52	2.334

(b) MWU group.

Table 6.3: Averaged interstage variability metrics and difference in gross efficiency between stage 1 and stage 3.

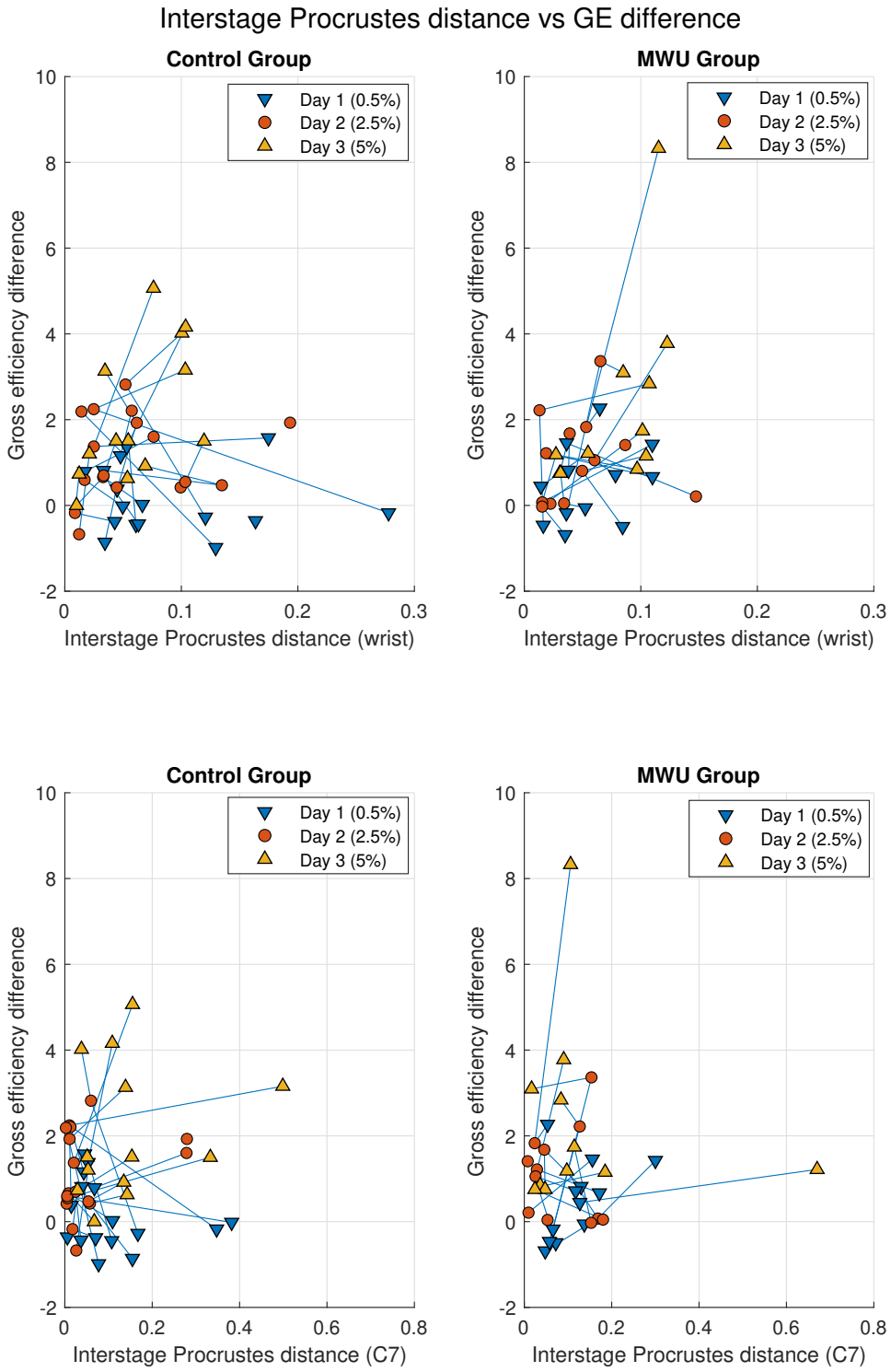


Figure 6.11: Interstage Procrustes distances and differences in gross efficiency from stage 1 to stage 3, labeled by experiment day. Linked points represent a single participant.

6.1.5 Propulsion Patterns

To investigate whether there were any significant differences in gross efficiency or energy expenditure based on the propulsion pattern of the wheelchair user, we will examine notched boxplots grouping the data from all stages by propulsion pattern. These are shown in Figure 6.12.

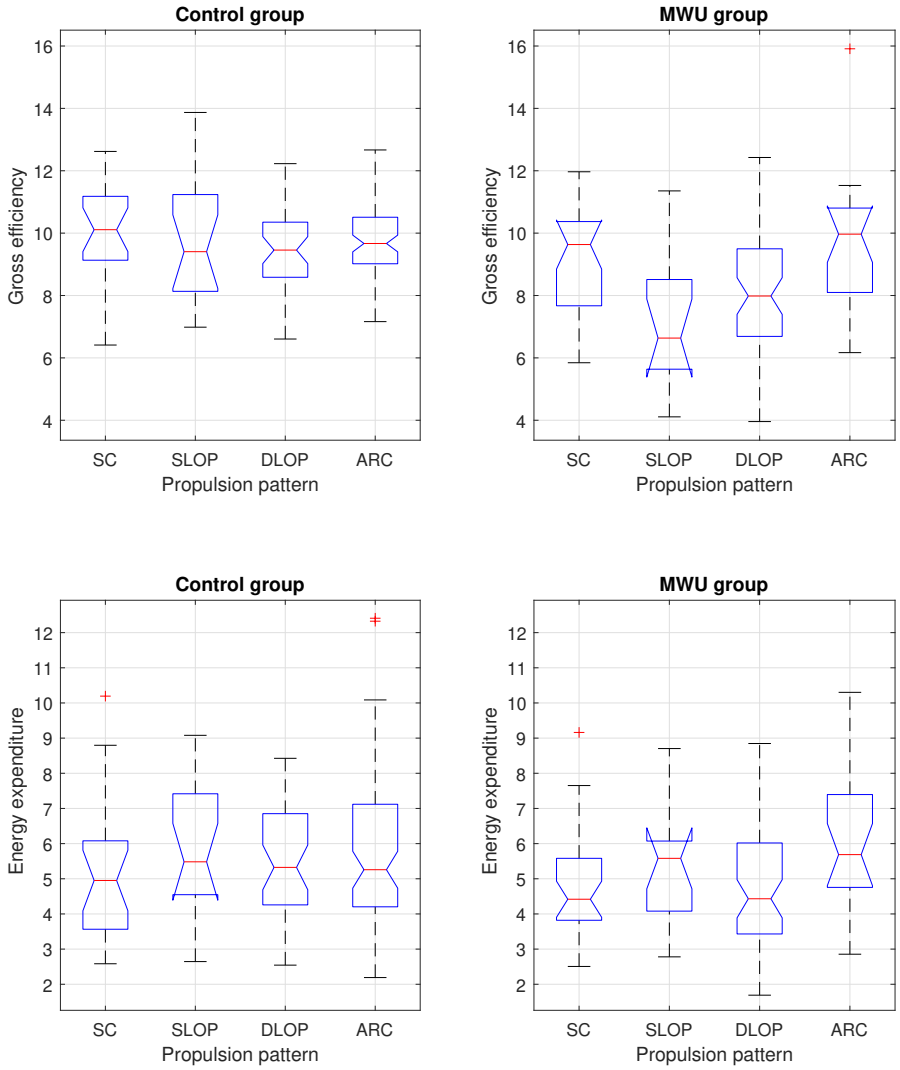


Figure 6.12: Boxplots of gross efficiency (top) and energy expenditure (bottom) grouped by propulsion pattern. The central mark represents the median, and the bottom and top edges respectively indicate the 25th and 75th percentiles. Points that are possible outliers are plotted using the "+" symbol. For independent data, two medians are considered significantly different at 5% significance level if the notches do not overlap.

In the control group, there seems to be no notable differences in either gross efficiency or the energy expenditure based on the propulsion patterns.

In the MWU group, the SC and ARC patterns are associated with higher gross efficiency than the other patterns, which may signify that there is a difference in the gross efficiency based on the propulsion pattern. However, there may be other factors that lead to these specific patterns. A possible reason may be that increased physical demand such as faster speeds or steeper inclines lead to specific propulsion patterns, which will be further investigated in Section 6.2.1. Regarding the energy expenditure in the MWU group, there are more differences between the propulsion patterns, and it seems that the ARC pattern is associated with higher energy expenditure than the others.

Since we included data from all stages for all participants, there were multiple contributions to the boxplots from each participant, which implies that the data is not independent. Based on the large notches, we can see that there is uncertainty in the true median value, which makes it hard to see whether there are small differences on a participant level. For instance, if there was a small systematic increase or decrease in one of the values on a participant level, this might be obscured by larger variations on the group level. A possible improvement could be to investigate how the values changes on a participant level. However, few participants utilize all of the propulsion classes, which makes this approach less viable as we wanted to investigate potential differences between the classes.

We should also note that in three of the boxplots we have notches that extend beyond the 25th or 75th percentiles, which may be an indication of a small sample size. Compared to other studies such as Slowik et al., 2015 which has gathered data from 170 MWUs, we are looking at data from only 20 participants in the control group and 15 in the MWU group. However, it should be mentioned that this thesis includes more data from each participant, as each participant goes through nine different sessions with different speeds and inclines during propulsion.

6.2 More on Propulsion Patterns

6.2.1 Effect of Speed and Incline on Propulsion Patterns

To see whether the propulsion patterns changed based on the speed and incline, we will illustrate in Figure 6.13 which patterns were used for each day-stage combination for each group.

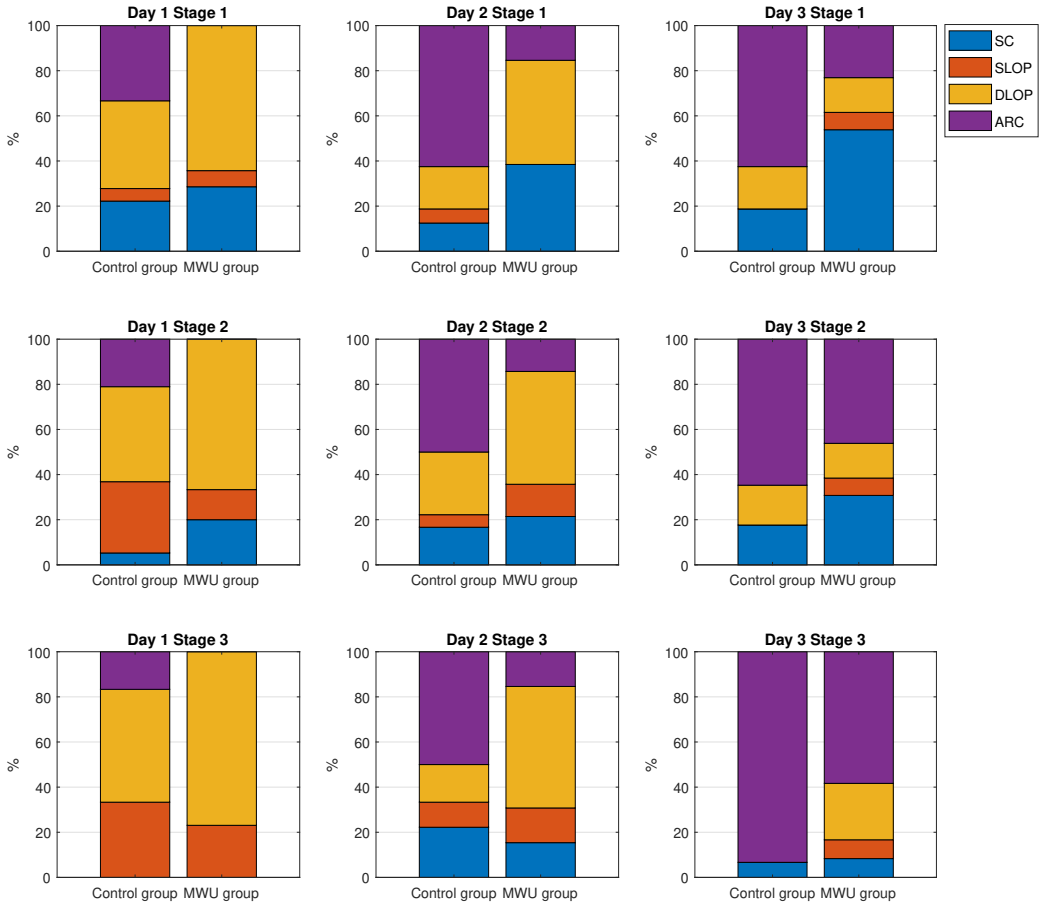


Figure 6.13: Propulsion patterns for each day-stage combination.

Starting with the control group, we can observe that the DLOP and SLOP patterns are more prominent in day 1, and the ARC pattern is more used in the higher incline days. For day 3, the ARC pattern is almost dominating, especially for the highest speed stage.

In the MWU group, we can also see that the DLOP pattern is more prominent in day 1, and the ARC pattern is not used. We can see the same trend that the ARC pattern is used more as the incline increases, though not as much as for the control group. The SC pattern is more common in this group.

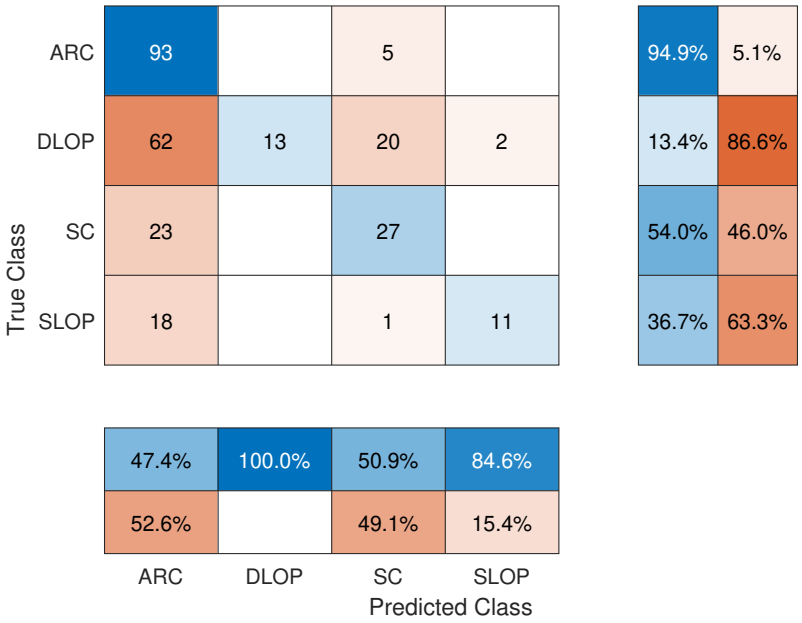
From both groups, we can observe that the patterns change both depending on the speed and the incline. The change based on speed can also be noticed based on the interstage variability shown in Section 6.1.4, which is significantly higher compared to intrastage variability.

6.2.2 Classification of Propulsion Patterns

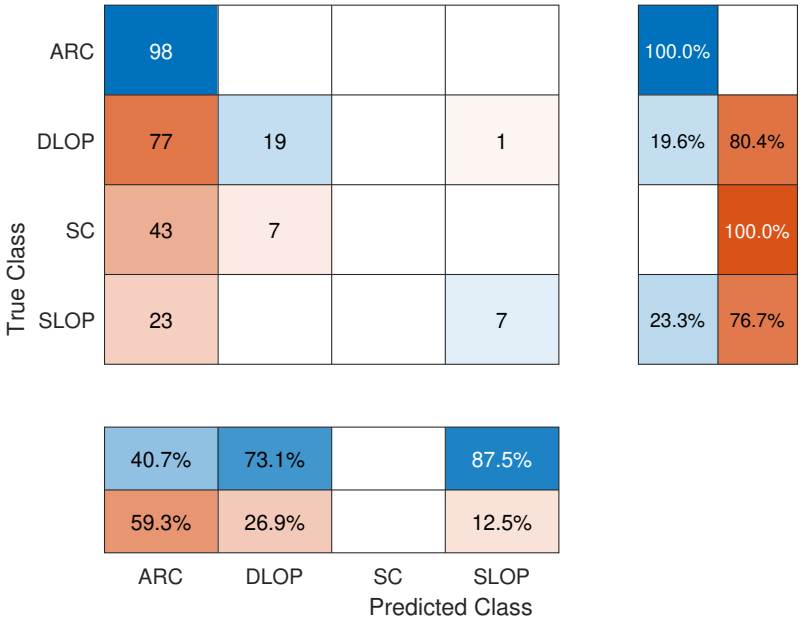
In Section 6.4 we will look at resulting motion trajectories from parsing IMU data through a neural network, but first we will look at another use-case for the generated trajectories. We will attempt to classify the propulsion pattern based on an averaged propulsion cycle using both Procrustes and DTW analysis as described in Section 5.5.

Figure 6.14a shows the confusion matrix from the Procrustes classifier. This method detects most of the ARC patterns, though it often mistakes other patterns as the ARC class. The DLOP class especially is often mistaken as the ARC class. This is not that surprising, as based on what we saw in Figure 3.1, the Procrustes method flipped the ARC pattern, resulting in DLOP being the most similar pattern compared to the other classes. However, in the cases where the classifier predicts that the pattern is in fact DLOP, it is correct in all cases. The classifier also tends to classify SC and SLOP patterns as ARC, though not to the same extent as with the DLOP pattern.

Figure 6.14b shows the confusion matrix from the DTW classifier. In this case as well, the classifier tends to classify most patterns as the ARC pattern. The classifier does not classify a single pattern as SC, and detects less of the SLOP patterns than the Procrustes classifier. From looking at the lower right of Figure 3.1, or in Figure A.6, we can see that the individual axes of the ARC pattern is not as complex as the other patterns, consisting of only two U-shapes. It might be easier for the DTW method to warp signals to best fit the ARC pattern, as it will not have to deal with looping trajectories. From the upper left of Figure A.6 we can see that the ARC and SC patterns have similar shapes, which is likely why the method is biased towards the ARC pattern.



(a) Confusion matrix from the Procrustes classifier.



(b) Confusion matrix from the DTW classifier.

Figure 6.14: Confusion matrices from pattern classifiers. The two columns on the right (from left to right) represents the positive predictive value and false discovery rate. The two rows on the bottom (from top to bottom) represents the sensitivity and false negative rate.

Neither of the classifiers had optimal performance, though the Procrustes classifier had most correct predictions of the two. Possible improvements to the method could be to include multiple samples of the pattern classes as a reference, as most trajectories do not match the archetypes perfectly.

6.3 Validity of IMU for Cycle Feature Extraction

Figure 6.15 shows the cycle features in the control group extracted from the wrist IMU plotted against cycle features from Qualisys data, before and after the synchronization method is applied. The unsynchronized data was extracted by assuming the first and second Qualisys window were exactly at respectively 60-90s and 180-210s relative to the IMU data. Figure 6.16 shows the same for the MWU group. There is a notable effect of the synchronization in both groups, and almost all points labeled with moderate or high intensity are on the reference line, which indicates perfect correlation between the IMU and Qualisys features. There are still some points, most labeled with low and moderate intensity, that are offset from the reference line. This is due to the cycle detecting method failing. The low-intensity outliers, generally from stage 1, are due to the method detecting too many cycles as the peaks in the acceleration signal are less discernible, which warrants a higher minimum time than 0.7s between detected extremal points. However, other outliers from higher intensity stages are due to the method detecting too few cycles due to the minimum time being too high. This indicates that the cycle detecting methods needs further tweaking, possibly with a threshold and minimum time between cycles that is adapted to the intensity.

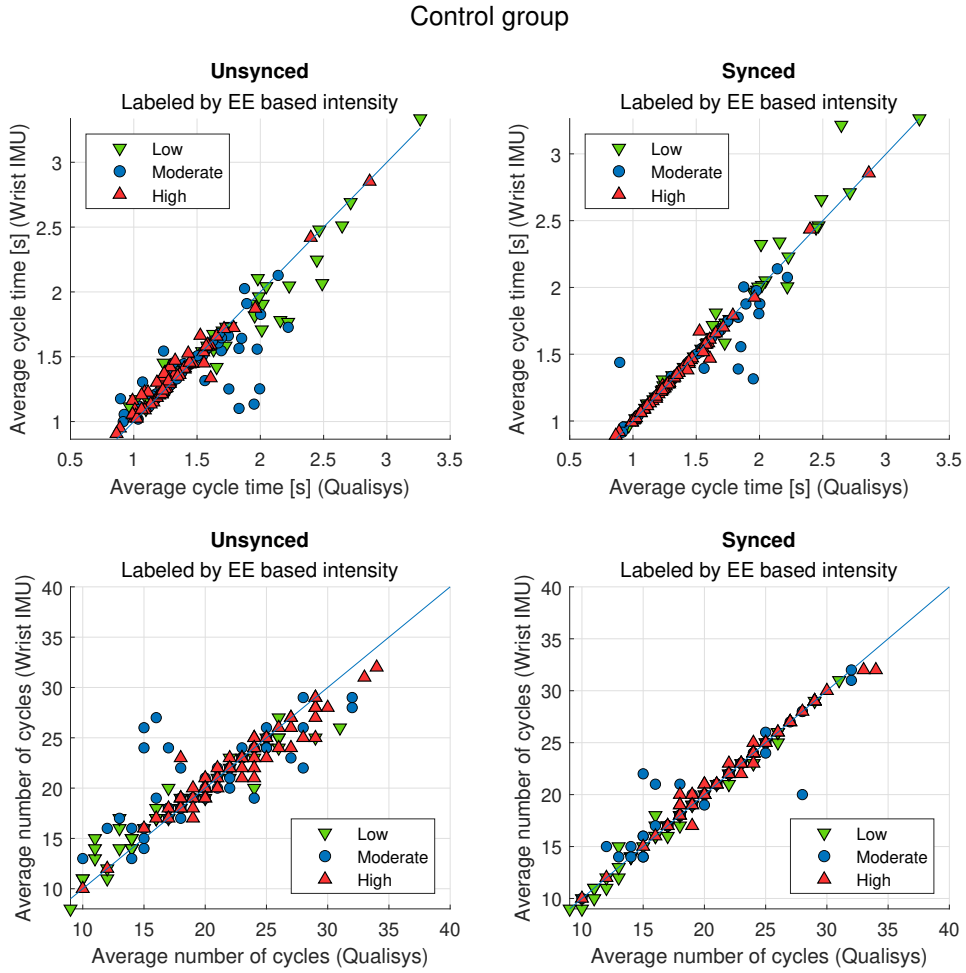


Figure 6.15: Cycle features from IMU and Qualisys data from the control group plotted against each other, before and after synchronization.

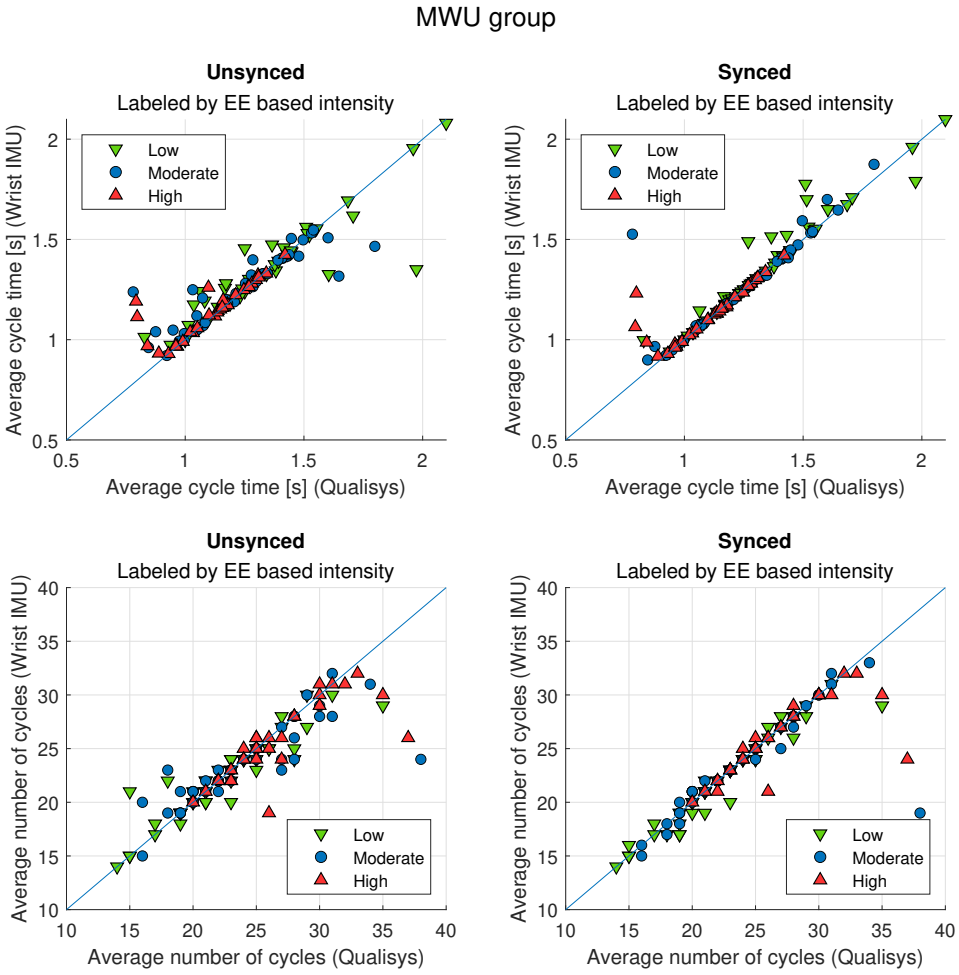


Figure 6.16: Cycle features from IMU and Qualisys data from the MWU group plotted against each other, before and after synchronization.

To illustrate the feature extraction and synchronization with a specific example, Figure 6.17 shows extracted extremal points plotted on top of the motion trajectories. Though both the unsynced and synced version produce the same cycle features, we can see how big a difference there is with just a few tenths of a second offset. The IMU X-maxima in the unsynced variant is spread out in the recovery phase of the propulsion cycle, whereas in the synced variant all are gathered just at the end of the push phase. While the difference isn't that important when just comparing the cycle features, the synchronization was essential when we later used the Qualisys data as a ground truth for generating trajectories based on IMU data.

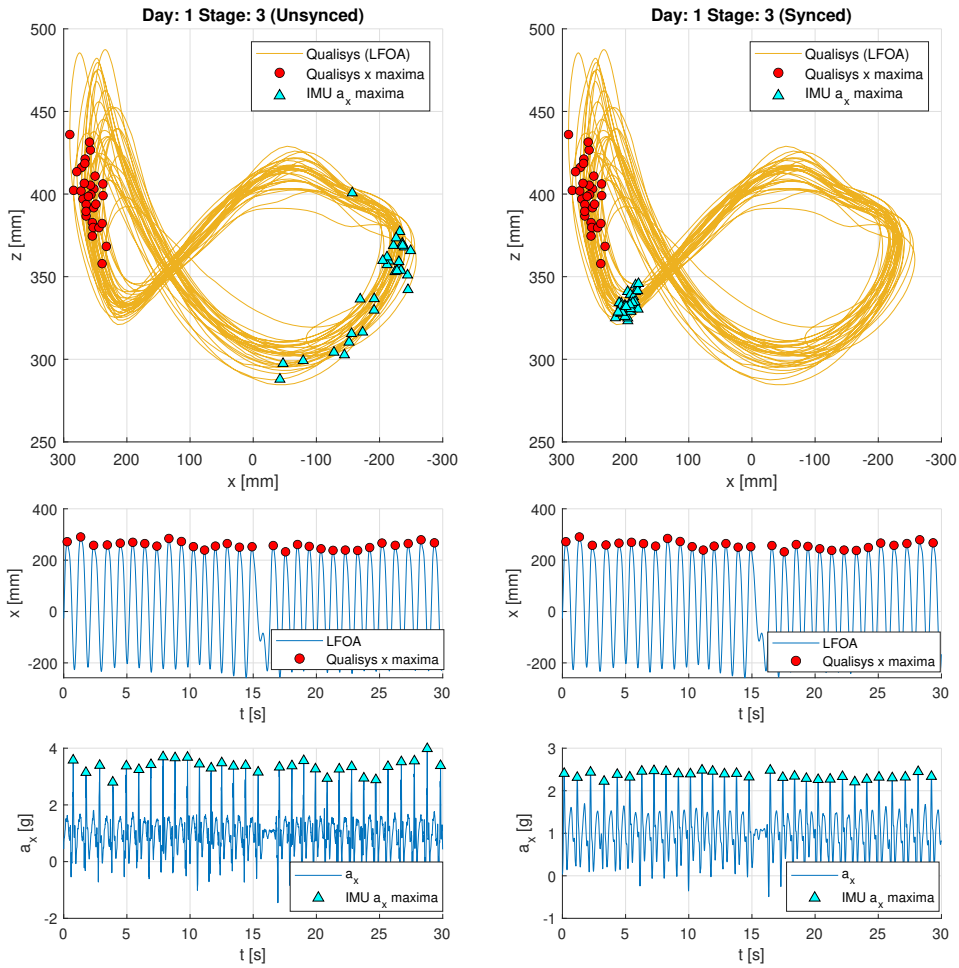


Figure 6.17: Motion signals and extracted extremal points, before and after synchronization. The motion signals are from the same session as shown in Figure 5.3.

Averaged cycle features from IMU- and Qualisys data per group for each day-stage combination are summarized in Table 6.4, and we can see that the IMU accurately captures the cycle features with small variations.

		Cycle Times		Number of Cycles	
		Qualisys	IMU	Qualisys	IMU
Day 1	Stage 1	1.82	1.82	16.50	16.44
	Stage 2	1.46	1.47	21.00	20.79
	Stage 3	1.28	1.30	24.72	24.17
Day 2	Stage 1	1.61	1.58	18.81	19.13
	Stage 2	1.46	1.46	20.56	20.78
	Stage 3	1.33	1.33	22.33	22.39
Day 3	Stage 1	1.80	1.81	16.63	17.31
	Stage 2	1.47	1.45	20.65	20.82
	Stage 3	1.25	1.25	23.87	23.87

(a) Control group.

		Cycle Times		Number of Cycles	
		Qualisys	IMU	Qualisys	IMU
Day 1	Stage 1	1.39	1.44	20.86	20.36
	Stage 2	1.21	1.23	24.60	24.13
	Stage 3	1.06	1.13	27.31	25.69
Day 2	Stage 1	1.35	1.37	21.92	21.54
	Stage 2	1.19	1.20	24.79	24.86
	Stage 3	1.08	1.11	27.31	26.54
Day 3	Stage 1	1.56	1.58	18.62	18.39
	Stage 2	1.24	1.24	23.92	23.92
	Stage 3	1.07	1.10	27.25	26.50

(b) MWU group.

Table 6.4: Comparison of averaged cycle features from Qualisys- and IMU data (after synchronization).

6.4 Reconstruction of Motion Trajectories Using a LSTM Neural Network

6.4.1 Performance Metrics

Performance metrics from training the neural network when partitioning out entire experiment days and stages are shown in Table 6.5a. The training- and validation performances are relatively similar for these partitions, though the gap is biggest from excluding day 1 and stage 2 in terms of RMSE. When using the trained network to predict on the test set, excluding day 1 gives a notably worse performance in all metrics. A possible reason for this is that in days 2 and 3 the overall propulsion patterns are different than in day 1. We saw in Section 6.2.1 that the ARC pattern is more prominent in days 2 and 3, and we can also note that especially for the control group there are few cases of the SC pattern in day 1 as well. This might bias the network towards propulsion patterns that are less used in day 1.

From excluding stages, the testing metrics are relatively similar with the exception of excluding stage 3, which gives a higher Procrustes distance than the rest. A possible reason for this is that the frequency of movement is higher in this stage, and the network might fail to reproduce correct features in these trajectories due to it being trained on data with a lower frequency.

There is not a linear relation between the RMSE or loss and the computed Procrustes distance, which can be seen by for instance comparing the training set metrics from excluding days 1 and 3. Although excluding day 3 gives a lower RMSE, it also gives the higher Procrustes distance.

Table 6.5b shows performance metrics from training the network on the control and MWU group separately, then using the other group as validation and test sets. Metrics from training on two

partitions of the validation and test sets for each group are shown, and we can see that though the network has the same training data for each partition, the performance metrics on the training sets are slightly perturbed in the control group. Possible reasons for this is due to the dropout layer of the network dropping out different data during each training, but may also be due to the fact that the network is trained on a GPU, which in MATLAB is not guaranteed to give a deterministic outcome (MathWorks, n.d.-b).

When partitioning out the groups, there is a relatively large gap in performance between the training and both the validation and test sets, which isn't unexpected as these sets consists of participants that are wholly unseen to the network. The validation performance is more similar between partitions when excluding the control group, and less so when excluding the MWU group. This is likely due to there only being five participants in the validation set when excluding the MWU group, making the set prone to variations in the data. A possible improvement could be to rather include all participants in the test set, but since we wanted to keep an equal amount of participants in the testing set for both groups, it was chosen to add the rest to the validation set. Another possibility would be to only have the validation set in the case where we exclude the control group, with 15 participants in each testing set. This was not done as we wanted to use the most of the available data while simultaneously facilitating comparison between the results from each group.

	Excluded Day/Stage:	Day 1	Day 2	Day 3	Stage 1	Stage 2	Stage 3
Training set	RMSE	0.606	0.626	0.556	0.579	0.641	0.577
	Loss	0.184	0.196	0.155	0.168	0.206	0.167
	Procrustes Distance	0.156	0.226	0.216	0.208	0.199	0.193
Validation set	RMSE	0.517	0.623	0.586	0.569	0.538	0.591
	Loss	0.134	0.194	0.172	0.162	0.145	0.175
	Procrustes Distance	0.172	0.250	0.243	0.221	0.209	0.204
Test set	RMSE	0.841	0.654	0.703	0.740	0.667	0.680
	Loss	0.371	0.225	0.259	0.284	0.233	0.245
	Procrustes Distance	0.365	0.209	0.168	0.197	0.197	0.255

(a) Performance metrics from training the neural network excluding entire days or stages.

	Excluded Group:	MWU (1)	MWU (2)	Control (1)	Control (2)
Training set	RMSE	0.572	0.573	0.612	0.634
	Loss	0.164	0.164	0.187	0.201
	Procrustes Distance	0.188	0.186	0.197	0.182
Validation set	RMSE	0.828	0.72	0.744	0.706
	Loss	0.367	0.261	0.281	0.250
	Procrustes Distance	0.252	0.283	0.272	0.257
Test set	RMSE	0.804	0.817	0.762	0.808
	Loss	0.341	0.354	0.303	0.350
	Procrustes Distance	0.324	0.284	0.229	0.246

(b) Performance metrics from training the neural network excluding participant groups. Metrics from two different partitions of validation and test sets for each group is shown.

Table 6.5: Neural network performance metrics.

6.4.2 Replicating Interstage Variability

To demonstrate the neural network, we will now look at a few examples of how it can be used to replicate the interstage variability feature. We will use Procrustes distance as the variability metric as it is easier to visualize the transformation between shapes using Procrustes analysis than looking at the individual axes using DTW analysis.

Figure 6.18 shows two examples of good replication of the interstage variability. This is based on

a threshold that the difference in variability between Qualisys and IMU trajectories must be less than 0.05. In the upper left of the figure we can see that the pattern changes from an obvious DLOP pattern in stage 1 to a more hybrid pattern in stage 3. In the upper right we can see the IMU-based trajectory, and for stage 1 the trajectory seems to take on the general shape of the Qualisys trajectory. For stage 3 however, the IMU trajectory is not as similar. However, the interstage variability computed is approximately equal. In the lower part the patterns are more similar to the ARC pattern, and the network is able to replicate the trajectories from both stages quite well.

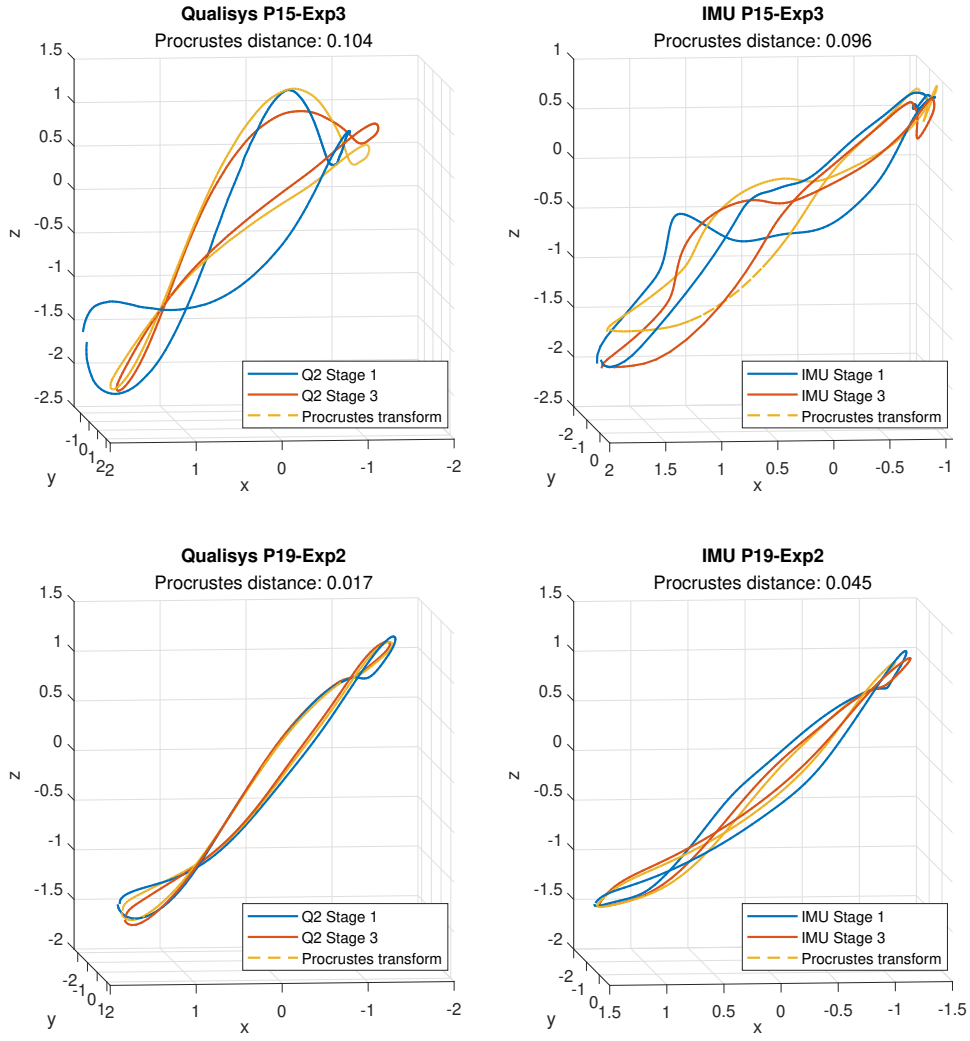


Figure 6.18: Good replication of interstage variability with neural network-generated trajectories. Each row represents an individual example.

Figure 6.19 shows two examples of bad replication of the interstage variability, based on a threshold that the difference in interstage variability is greater than 0.2. In the upper part of the figure we can see that the network replicates the trajectory from stage 1 quite well, but misses on stage 3. In the stage 3 trajectories, the Qualisys variant seems to be an SC pattern, though the IMU produces a trajectory most similar to the DLOP pattern. This is somewhat similar to what we saw for stage 3 in the upper part of Figure 6.18, and might indicate that we lack a certain richness in the trajectories of the training data. From the Qualisys trajectories in the lower part of Figure 6.19, we can see that the pattern changes from a SC pattern in stage 1 to a DLOP pattern in stage

3. The IMU trajectories from stage 1 are somewhat able to recreate the SC shape, though the trajectory is much more rugged and may be more similar to a DLOP pattern. For stage 3, the IMU trajectory is more similar to the Qualisys trajectory, though there is an indent in the leftmost part which was likely the cause of the high Procrustes distance.

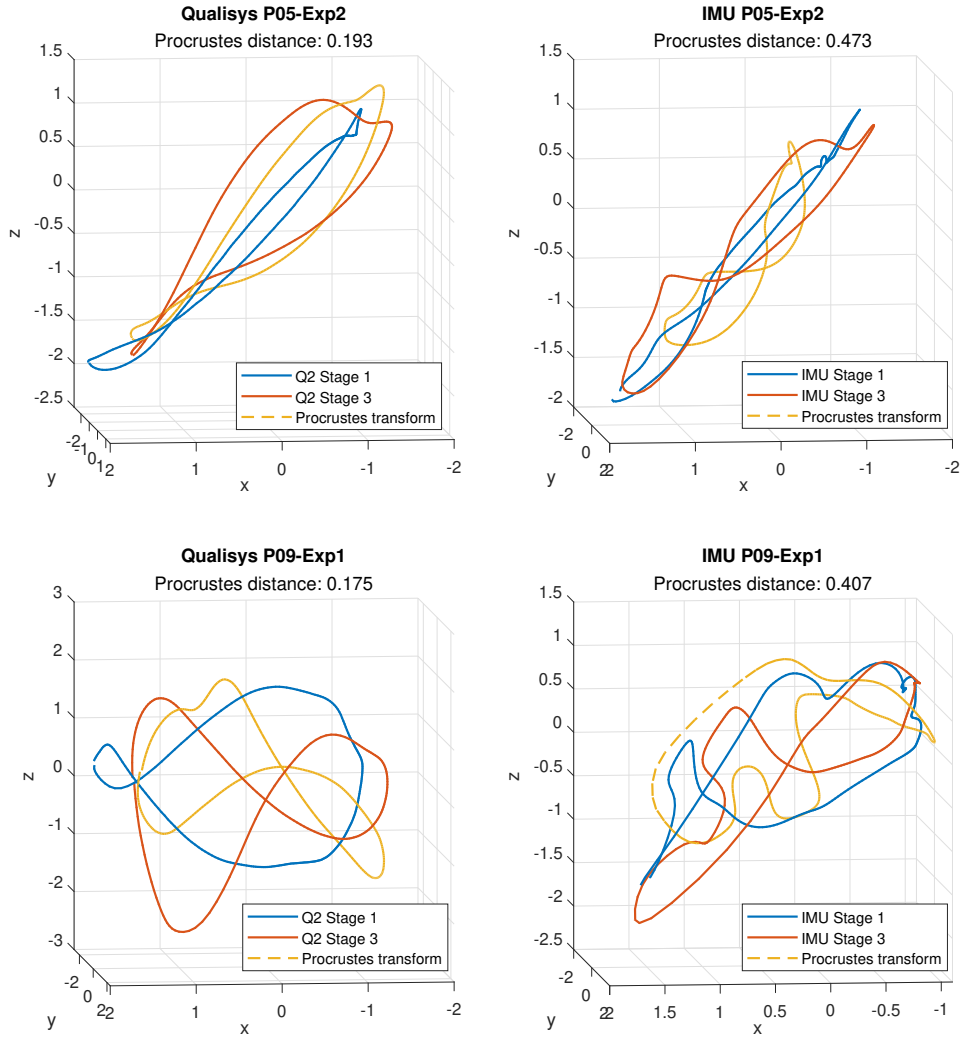


Figure 6.19: Bad replication of interstage variability with neural network-generated trajectories. Each row represents an individual example.

Trajectories from the neural network are generally more rugged, and it is better at correctly reproducing patterns in the ARC class, which is generally less complex than the other patterns. More examples of both good and bad replication of the interstage variability are added in Appendix A.5.

7 Discussion

7.1 Relation Between Proposed Features and Physiological Values

7.1.1 Cycle Features

In Section 6.1.1 we explored the relation between all of the proposed features and physiological values using PCA as a tool. From this, energy expenditure showed positive correlation with the number of cycles, reflecting the higher energy requirement with increased movement. We should note, however, that the cycle times is not a novel feature, and other studies such as Spaepen et al., 1996 has investigated propulsion cycle times in relation to energy expenditure. The cycle features were investigated further in Section 6.1.2, where we found that the energy expenditure was higher when the number of cycles were high and the cycle times were low. Regarding gross efficiency, it was harder to do any separation based on the cycle features, though there were signs of a trend that the gross efficiency was higher for stages 2 and 3. Ettema and Lorås, 2009 describes a curved work rate-gross efficiency relationship in cycling, where the gross efficiency increases with the work rate (speed), then flattens out after a certain threshold. This is likely the cause of the differences in gross efficiency between the first and the other stages, where for stages 2 and 3 the curve has stabilized. This increase was more obvious in the control group than in the MWU group, which seems to indicate that the higher speed stages forces the wheelchair user to take on a more efficient movement, even though they have little experience with wheelchair propulsion. Since we are able to a certain extent determine which stage data comes from based on the cycle features, these could be useful in estimating the energy expenditure and gross efficiency of the user.

7.1.2 Intrastage Variability

The intrastage variability feature allowed us to quantify how much movement patterns changed over time when the speed and incline of the propulsion movement were constant. In Section 6.1.3 we examined correlations between the intrastage variability and physiological values, and found a slight anticorrelation between the intrastage variability and the gross efficiency. This was more obvious in the control group than in the MWU group. While it was hard to do any separation of either the experiment days or stages based on the feature, there was a trend that the sessions with the highest gross efficiency had small intrastage variability. This was more clear for the DTW distances, which as seen in Figure 6.9 showed that there was a slight upward trend in the gross efficiency for both the control and MWU groups as the DTW distances got lower. This seems to support the initial hypothesis, that low variability might imply more efficient propulsion. However, there were almost equally many points with low variability that were associated with low gross efficiency, so it is likely other factors than the intrastage variability that are more closely related to the gross efficiency.

Based on these findings, it is unclear whether a biomechanical feedback tool should encourage propulsion movement with little or high variability in order to achieve higher gross efficiency. On the other hand, the feature may have other applications; Sosnoff et al., 2015 found that movement with higher spatial variability was correlated with shoulder pain in the individual, which may warrant the feedback tool to recommend movement with little variability. However, it may be uncertain whether the variability is the cause of the shoulder pain, or if the shoulder pain is the cause of the variability. Hamill et al., 2012 suggests that shoulder pain can develop from either too little or too much motor variability, indicating that encouraging a specific propulsion variability might not be viable.

7.1.3 Interstage Variability

The interstage variability feature presented a metric of how much the movement patterns changed with increasing speed, and thus increasing physical demand of the exercise. This metric was compared to whether there was a corresponding increase or decrease in gross efficiency, with the initial

idea that this variability should be larger than the intrastage variability. Comparison between Tables 6.2 and 6.3 shows that the interstage variability is indeed higher for both the control and MWU groups. The interstage variability was at its lowest in the 2.5% incline day, which may be due to the relation between speed and incline. From Table 4.3 we can see that for the 0.5% incline day, the speed is increased by 4 km/h from the first to the last stage. For the other days, the increase is only 2 km/h. From the averaged physiological values in Table 6.2, we can see that the energy expenditure for stage 3 is lower for the 0.5%- and 2.5% incline days in both groups. Given this, a possible explanation for high interstage variability in the 0.5% incline day might be that the movement is less physically taxing, which leaves the wheelchair user with more freedom of movement. In the 2.5% incline day, we see less variability, likely due to the movement becoming more demanding and forcing the wheelchair user into a more consistent pattern. Finally, for the 5% incline day, the variability is higher. This may be due to the 2 km/h increase in speed being more significant for the steeper incline, which makes it hard to keep a consistent propulsion pattern across stages even though the propulsion is physically demanding.

Due to the fact that the interstage variability do not monotonically decrease or increase with the incline, it may be difficult to use this as a feature to estimate how the gross efficiency changes without any prior information. We saw that the gross efficiency increased more between stages when the incline was steeper, without seeing a similar trend in the interstage variability. A possible solution could be to estimate the incline, which in combination with the variability could tell something about how the gross efficiency changes. However, in this case we would also need to estimate the speed as the speed was increased depending on the incline.

7.1.4 Propulsion Patterns

In Section 6.1.5 we examined the relation between the propulsion patterns and both gross efficiency and energy expenditure. We found that the gross efficiency of the SC and ARC patterns was higher than for the other patterns in the MWU group, which may suggest that these patterns should be favoured. This is partly supported by Slowik et al., 2016, which found the SC pattern to produce the most favourable levels of upper extremity demand. However, as we saw in Figure 6.13, the propulsion patterns seems to be a product of the speed and incline of treadmill, rather than just an individual technique. The reason that we see higher gross efficiency in the SC and ARC patterns is likely be due to these patterns being more used in days 2 and 3, which is more demanding than the first. We see a shift towards the ARC pattern in both groups as the incline increases, which is likely why the ARC pattern is associated with higher energy expenditure in Figure 6.12. The shift towards the ARC pattern is in line with the observations made in Slowik et al., 2015, which further suggests that there is not one optimal pattern for all propulsion tasks, but rather that it should be adapted depending on the required physical demand.

7.2 Validity of Features from IMU Data

In order to be able to use the analysed features towards a biomechanical feedback device, a primary goal was to be able to reproduce them with equipment that is available outside of a lab setting.

In Section 6.3 we showed that the cycle features can in most cases be reliably extracted from acceleration signals, with the main exception being during low intensity propulsion. This can likely be improved by tuning the cycle detection method. However, in a daily-life setting another extraction might be favoured. Allseits et al., 2017 proposed a real-time method called *Noise-Zero Crossing* as an alternative to an extremal point-based detection method in gait analysis, which could possibly be adapted to wheelchair propulsion to improve detection in the low-intensity stages.

To be able to reproduce the variability features, we examined generating trajectories from IMU accelerations using a LSTM neural network in Section 6.4. The network was trained multiple times with different partitions of the training data in order to demonstrate the sensitivity of the network to different training data. Using the Procrustes distances as a performance metric, these were generally higher than 0.2 on the test set. In Figure 3.1 we saw that when comparing the archetypal propulsion patterns to each other, a Procrustes distance of just 0.1 could indicate a totally different

propulsion pattern, which indicates that there is much room for improvement in the reconstruction. However, we also saw that the network was able to reproduce the interstage variability in specific cases. The testing metric was computed by comparing the IMU trajectories to the original Qualisys trajectories, but the variability features are computed by comparing trajectories from each source to each other. If the variability features from each source were similar enough, the performance gap would not be as critical. However, we saw in Figure 6.19 that the network may output shapes that have different characteristics than the original ones, which might indicate that the network is biased to certain propulsion patterns such as the DLOP pattern. This is likely due to the DLOP pattern being one of the most used patterns in the dataset, in addition to the ARC pattern. In order for the network to be useful in a feedback device, it should at least be able to discern the propulsion patterns from each other.

A specific improvement to the reconstruction of the trajectories would be to include data relating to the orientation of the IMU as an input in the neural network, such as either gyroscope data or estimated quaternions. Studies such as Konak et al., 2020 and Lin et al., 2022 that achieve better motion tracking accuracy than we did in this thesis included gyroscope data, and the latter also included data from a magnetometer. Acceleration signals from different propulsion techniques has similar structure due to the largest accelerations, at least in the x-direction, corresponding to the same part of the propulsion cycle, and the rest of the signal is what separates them. There may not be sufficient information in these parts of the signal to reliably separate the different propulsion techniques, which can possibly be improved by adding another signal representing the movement.

7.3 Methodological Considerations

The extraction of cycles features relied on a minimum time of 0.7s between cycles. Most human motion is in the frequency range of 0.3 to 3.5Hz (Nightingale et al., 2014), which implies that the specified minimum time may be too strict. However, in the lab setting we were able to manually identify the specific cases where this was the case, and these were considered too few to warrant further tuning. It was considered a trade-off between being able to reliably detect cycles in low intensity and higher intensity propulsion, and the minimum time of 0.7s seemed to be a middle ground.

In the lab setting, we expected the IMU data to contain signals only corresponding to propulsion cycles with known speeds and inclines, and were able to tune the method accordingly. Outside of the lab, wheelchair users will move their arms in movements not necessarily including propulsion. This warrants that the feedback device is either able to reliably detect when propulsion occurs, or it needs to have a setting explicitly for propulsion. Additionally, it should be expected that there is variations in the speed, incline, and also direction of propulsion within short time-windows. These are factors that need to be considered if any of the features are to be used in a predictive model.

The labeling of propulsion patterns in this thesis was done manually by one person, which may have resulted in some mislabeling especially for more hybrid patterns. Due to the scope of this thesis being more exploratory, this was considered acceptable. Further studies, however, should utilize other methods for labeling such as a consensus of multiple raters as in Koontz et al., 2009, or more machine learning based methods such as the fuzzy clustering method presented in Aissaoui and Desroches, 2008.

In Section 6.2.2 we examined using Procrustes- and DTW analysis for classification of propulsion patterns, which both achieved sub-par performance compared to other methods such as those presented in Section 2.2. The performance of both classifiers can likely be improved by adding more examples as representations of the propulsion classes, but there might be other tools that are more viable. Using an approach similar to French et al., 2008 or Rindal et al., 2018, better accuracy can likely be achieved using only IMU data and neural networks, without having to generate the trajectories first.

7.4 Future Work

Future work should aim to further investigate the effect of variability in propulsion patterns. We have proposed novel methods of quantifying variability, but should be wary of encouraging a specific variability from the wheelchair user without knowing the exact physical effects of the variability.

In both the intrastage- and interstage variability metrics, there were more variability in the movement of the C7 marker than the wrist marker. This is possibly due to the upper body being less constrained during wheelchair propulsion, as the handrim of the wheel guides the hand for a significant part of each propulsion cycle. An interesting approach would be to remove the part of the wrist trajectory corresponding to when the hand is in contact with the wheel, then seeing whether there are more notable differences between propulsion patterns.

When the data from the whole MWU group is available, the new participants should be included in the validation set and the metrics in Table 6.5b should be updated to see investigate whether the small validation sets were the cause of the gap in performance between partitions when training exclusively on the control group.

8 Conclusion

This thesis examined the relation between frequency- and variability-based features in manual wheelchair propulsion and their correlation with physiological values, and how to reproduce these features from equipment such as IMUs that are available outside of a lab setting.

By analyzing the frequency-based features, the thesis revealed a positive correlation between the number of cycles in a given time window and energy expenditure. Results indicate that propulsion at higher speeds results in lower propulsion cycle times, which is correlated with a higher gross efficiency. The thesis demonstrated that the cycle features can be reliably extracted from IMU data, with some adaptations needed in order to perfectly extract them from very low or high intensity propulsion.

Regarding the variability-based features, drawing specific conclusions about their relation to physiological values proved challenging. Results indicated that the propulsion pattern was adapted by the wheelchair user as a product of the speed and incline of the propulsion movement. It is also possible to note a shift of the trajectories towards an arcing pattern as physical demand increases. Moreover, the interstage variability was measured to be generally higher than the intrastage one, another phenomenon which reflects this shift. Finally, our method of demonstrating how these features can be extracted outside of a lab setting relied on a neural network for generating motion trajectories from IMU data. This approach was moderately successful in being able to accurately reproduce the original, motion-tracked trajectories.

We envision that future research should explore alternative applications of the variability-based features and focus on improving their extraction method.

References

Aissaoui, R., & Desroches, G. (2008). Stroke pattern classification during manual wheelchair propulsion in the elderly using fuzzy clustering. *Journal of Biomechanics*, *41*(11), 2438–2445. https://doi.org/https://doi.org/10.1016/j.jbiomech.2008.05.020

Allseits, E., Lučarević, J., Gailey, R., Agrawal, V., Gaunaurd, I., & Bennett, C. (2017). The development and concurrent validity of a real-time algorithm for temporal gait analysis using inertial measurement units. *Journal of Biomechanics*, *55*, 27–33. https://doi.org/10.1016/j.jbiomech.2017.02.016

Boninger, M. L., Souza, A. L., Cooper, R. A., Fitzgerald, S. G., Koontz, A. M., & Fay, B. T. (2002). Propulsion patterns and pushrim biomechanics in manual wheelchair propulsion. *Archives of Physical Medicine and Rehabilitation*, *83*(5), 718–723. https://doi.org/https://doi.org/10.1053/apmr.2002.32455

Collins, E. G., Gater, D., Kiratli, J., Butler, J., Hanson, K., & Langbein, W. E. (2010). Energy cost of physical activities in persons with spinal cord injury. *Medicine and Science in Sports and Exercise*, *42*(4), 691–700. https://doi.org/10.1249/mss.0b013e3181bb902f

Ettema, G., & Lorås, H. W. (2009). Efficiency in cycling: A review. *European Journal of Applied Physiology*, *106*(1), 1–14. https://doi.org/10.1007/s00421-009-1008-7

Filippeschi, A., Schmitz, N., Miezal, M., Bleser, G., Ruffaldi, E., & Stricker, D. (2017). Survey of motion tracking methods based on inertial sensors: A focus on upper limb human motion. *Sensors*, *17*(6). https://doi.org/10.3390/s17061257

French, B., Smailagic, A., Siewiorek, D., Ambur, V., & Tyamagundlu, D. (2008). Classifying wheelchair propulsion patterns with a wrist mounted accelerometer. *BODYNETS 2008 - 3rd International ICST Conference on Body Area Networks*, 20. https://doi.org/10.4108/ICST.BODYNETS2008.2963

Gaesser, G., & Brooks, G. (1975). Muscular efficiency during steady-rate exercise: Effects of speed and work rate. *Journal of applied physiology*, *38*, 1132–9. https://doi.org/10.1152/jappl.1975.38.6.1132

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, *40*. Retrieved 1st December 2022, from https://doi.org/10.1007/BF02291478

Hagströmer, M., Oja, P., & Sjöström, M. (2006). The international physical activity questionnaire (ipaq): A study of concurrent and construct validity. *Public Health Nutrition*, *9*(6), 755–762. https://doi.org/10.1079/PHN2005898

Hamill, J., Palmer, C., & Emmerik, R. E. A. V. (2012). Coordinative variability and overuse injury. *Sports Medicine, Arthroscopy, Rehabilitation, Therapy and Technology*, *4*(1). https://doi.org/10.1186/1758-2555-4-45

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, Y., Kaufmann, M., Aksan, E., Black, M. J., Hilliges, O., & Pons-Moll, G. (2018). Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time. *37*(6). https://doi.org/10.1145/3272127.3275108

Jetté, M., Sidney, K., & Blümchen, G. (1990). Metabolic equivalents (mets) in exercise testing, exercise prescription, and evaluation of functional capacity. *Clinical Cardiology*, *13*(8), 555–565. https://doi.org/https://doi.org/10.1002/clc.4960130809

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kirshblum, S. C., Waring, W., Biering-Sorensen, F., Burns, S. P., Johansen, M., Schmidt-Read, M., Donovan, W., Graves, D. E., Jha, A., Jones, L., Mulcahey, M. J., & Krassioukov, A. (2011). Reference for the 2011 revision of the international standards for neurological classification of spinal cord injury [PMID: 22330109]. *The Journal of Spinal Cord Medicine*, *34*(6), 547–554. https://doi.org/10.1179/107902611X13186000420242

Konak, O., Wegner, P., & Arnrich, B. (2020). Imu-based movement trajectory heatmaps for human activity recognition. *Sensors*, *20*(24). https://doi.org/10.3390/s20247179

Koontz, A. M., Roche, B. M., Collinger, J. L., Cooper, R. A., & Boninger, M. L. (2009). Manual wheelchair propulsion patterns on natural surfaces during start-up propulsion. *Archives of*

Physical Medicine and Rehabilitation, *90*(11), 1916–1923. https://doi.org/https://doi.org/10.1016/j.apmr.2009.05.022

Lin, B.-S., Lee, I.-J., Wang, S.-P., Chen, J.-L., & Lin, B.-S. (2022). Residual neural network and long short-term memory–based algorithm for estimating the motion trajectory of inertial measurement units. *IEEE Sensors Journal*, *22*(7), 6910–6919. https://doi.org/10.1109/JSEN.2022.3153398

Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (2015). Smpl: A skinned multi-person linear model. *34*(6). https://doi.org/10.1145/2816795.2818013

MathWorks. (n.d.-a). *Matlab long short-term memory neural networks* (Version R2022b). Retrieved 26th April 2023, from https://se.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html

MathWorks. (n.d.-b). Scale up deep learning in parallel, on gpus, and in the cloud. Retrieved 19th May 2023, from https://se.mathworks.com/help/deeplearning/ug/scale-up-deep-learning-in-parallel-on-gpus-and-in-the-cloud.html

Mercer, J. L., Boninger, M., Koontz, A., Ren, D., Dyson-Hudson, T., & Cooper, R. (2006). Shoulder joint kinetics and pathology in manual wheelchair users. *Clinical Biomechanics*, *21*(8), 781–789. https://doi.org/https://doi.org/10.1016/j.clinbiomech.2006.04.010

Nightingale, T. E., Rouse, P. C., Thompson, D., & Bilzon, J. L. J. (2017). Measurement of physical activity and energy expenditure in wheelchair users: Methods, considerations and future directions. *Sports Medicine - Open*, *3*(1). https://doi.org/10.1186/s40798-017-0077-0

Nightingale, T. E., Walhim, J.-P., Thompson, D., & Bilzon, J. L. J. (2014). Predicting physical activity energy expenditure in manual wheelchair users. *Medicine and Science in Sports and Exercise*, *46*(9), 1849–1858. https://doi.org/10.1249/mss.0000000000000291

Rindal, O. M. H., Seeberg, T. M., Tjønnås, J., Haugnes, P., & Sandbakk, Ø. (2018). Automatic classification of sub-techniques in classical cross-country skiing using a machine learning algorithm on micro-sensor data. *Sensors*, *18*(1). https://doi.org/10.3390/s18010075

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *26*(1), 43–49. https://doi.org/10.1109/TASSP.1978.1163055

Silverstein, B. A., Fine, L. J., & Armstrong, T. J. (1987). Occupational factors and carpal tunnel syndrome. *American Journal of Industrial Medicine*, *11*(3), 343–358. https://doi.org/10.1002/ajim.4700110310

Slowik, J. S., Requejo, P. S., Mulroy, S. J., & Neptune, R. R. (2015). The influence of speed and grade on wheelchair propulsion hand pattern. *Clinical Biomechanics*, *30*(9), 927–932. https://doi.org/https://doi.org/10.1016/j.clinbiomech.2015.07.007

Slowik, J. S., Requejo, P. S., Mulroy, S. J., & Neptune, R. R. (2016). The influence of wheelchair propulsion hand pattern on upper extremity muscle power and stress. *Journal of Biomechanics*, *49*(9), 1554–1561. https://doi.org/10.1016/j.jbiomech.2016.03.031

Sola, J., & Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on Nuclear Science*, *44*(3), 1464–1468. https://doi.org/10.1109/23.589532

Sosnoff, J. J., Rice, I. M., Hsiao-Wecksler, E. T., Hsu, I. M. K., Jayaraman, C., & Moon, Y. (2015). Variability in wheelchair propulsion: A new window into an old problem. *Frontiers in Bioengineering and Biotechnology*, *3*. https://doi.org/10.3389/fbioe.2015.00105

Spaepen, A. J., Vanlandewijck, Y. C., & Lysens, R. J. (1996). Relationship between energy expenditure and muscular activity patterns in handrim wheelchair propulsion. *International Journal of Industrial Ergonomics*, *17*(2), 163–173. https://doi.org/10.1016/0169-8141(95)00047-x

Tsang, K., Hiremath, S. V., Crytzer, T. M., Dicianno, B. E., & Ding, D. (2016). Validity of activity monitors in wheelchair users: A systematic review. *Journal of Rehabilitation Research and Development*, *53*(6), 641–658. https://doi.org/10.1682/jrrd.2016.01.0006

Weir, J. B. d. V. (1949). New methods for calculating metabolic rate with special reference to protein metabolism. *The Journal of Physiology*, *109*(1-2), 1–9. https://doi.org/10.1113/jphysiol.1949.sp004363

Wilby, M. L. (2019). Physical mobility impairment and risk for cardiovascular disease. *Health Equity*, *3*(1), 527–531. https://doi.org/10.1089/heq.2019.0065

Appendix

A Plots

A.1 Procrustes and DTW Examples

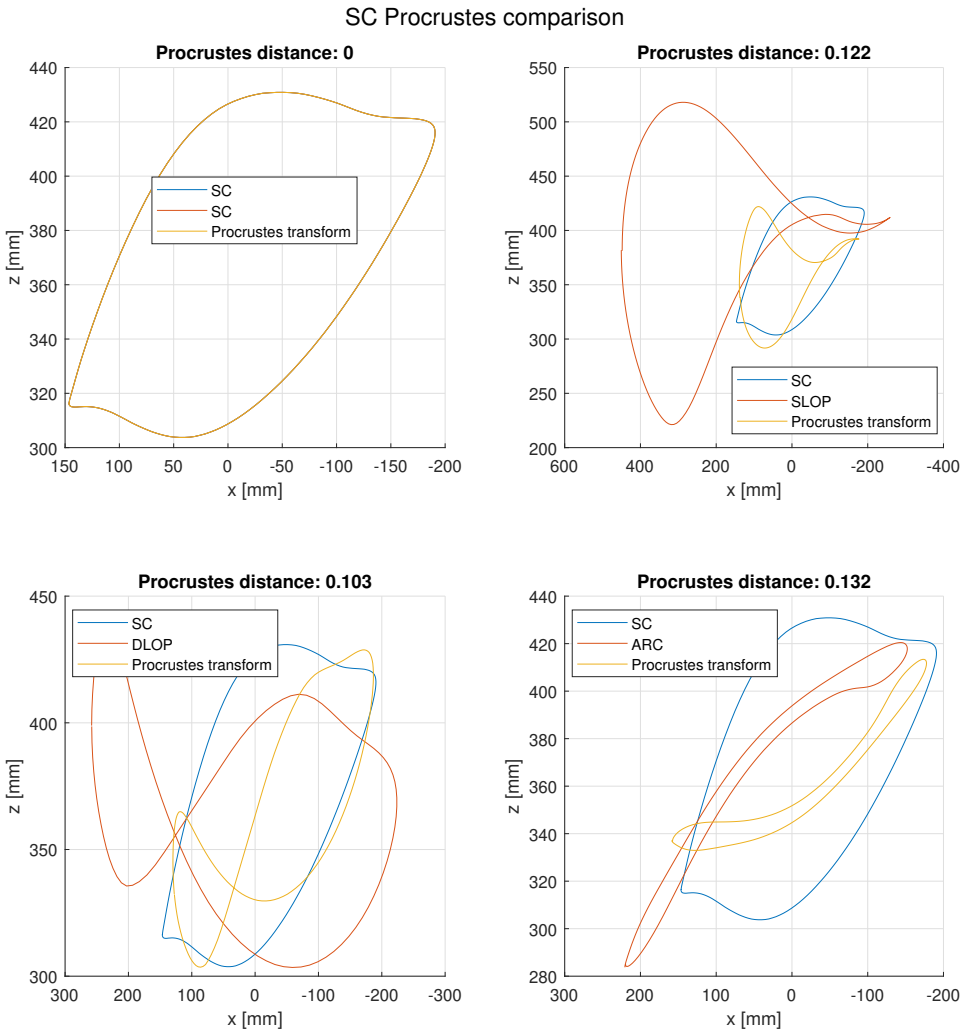


Figure A.1: Comparison of propulsion patterns using Procrustes analysis. Here, all shapes are compared to the SC pattern.

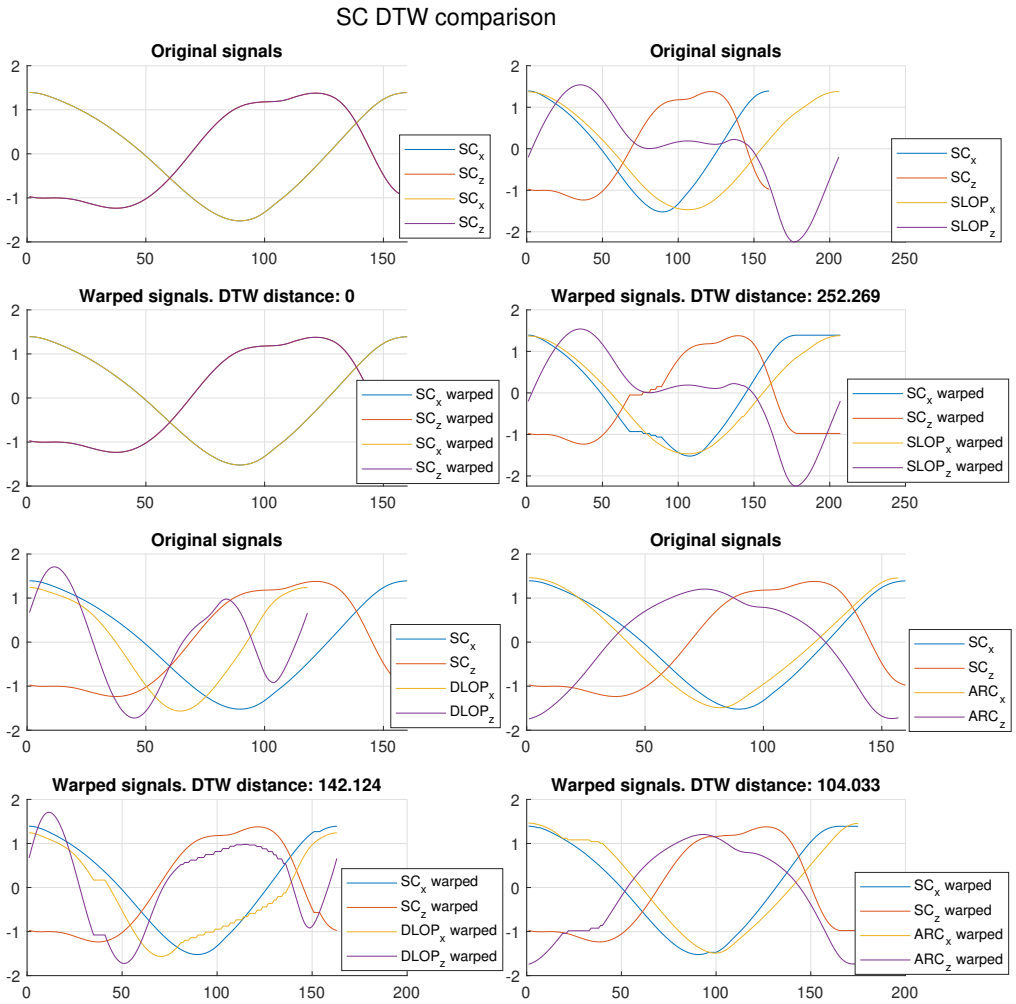


Figure A.2: Comparison of propulsion patterns using DTW analysis. Here, all shapes are compared to the SC pattern.

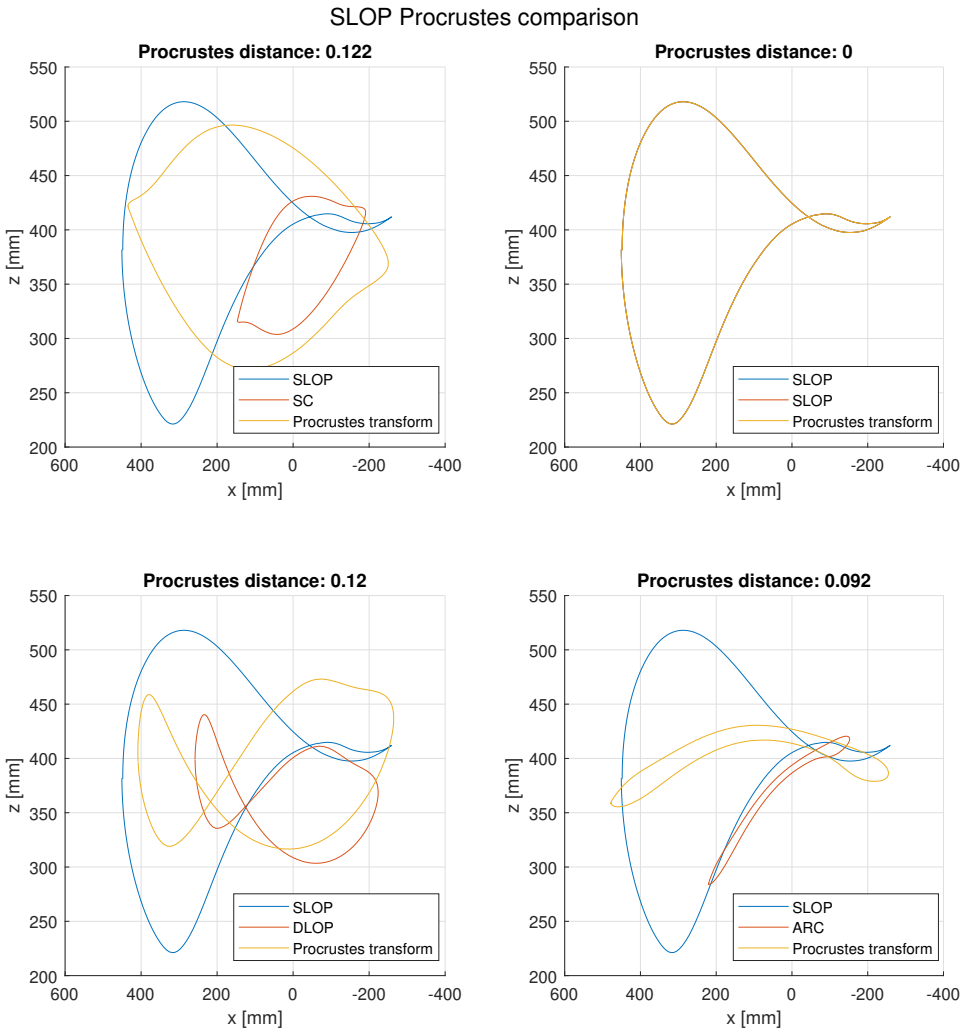


Figure A.3: Comparison of propulsion patterns using Procrustes analysis. Here, all shapes are compared to the SLOP pattern.

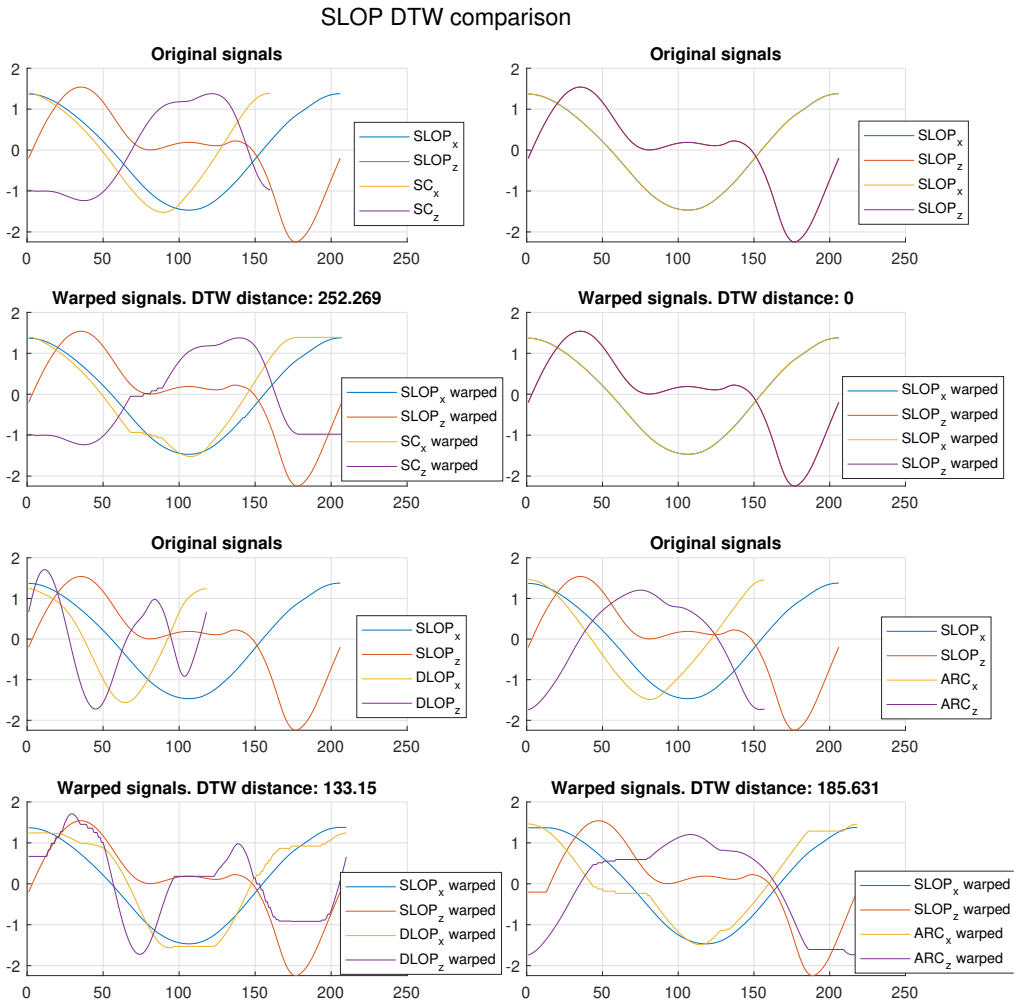


Figure A.4: Comparison of propulsion patterns using DTW analysis. Here, all shapes are compared to the SLOP pattern.

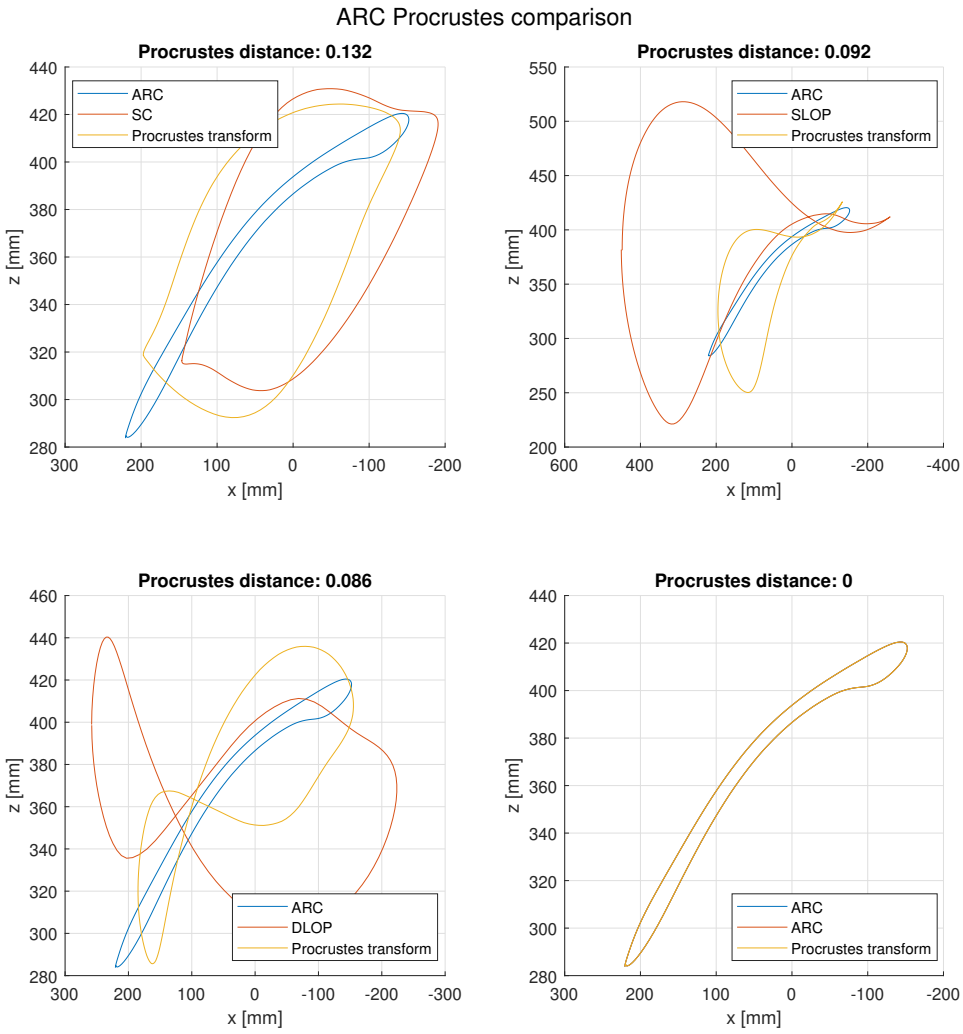


Figure A.5: Comparison of propulsion patterns using Procrustes analysis. Here, all shapes are compared to the ARC pattern.

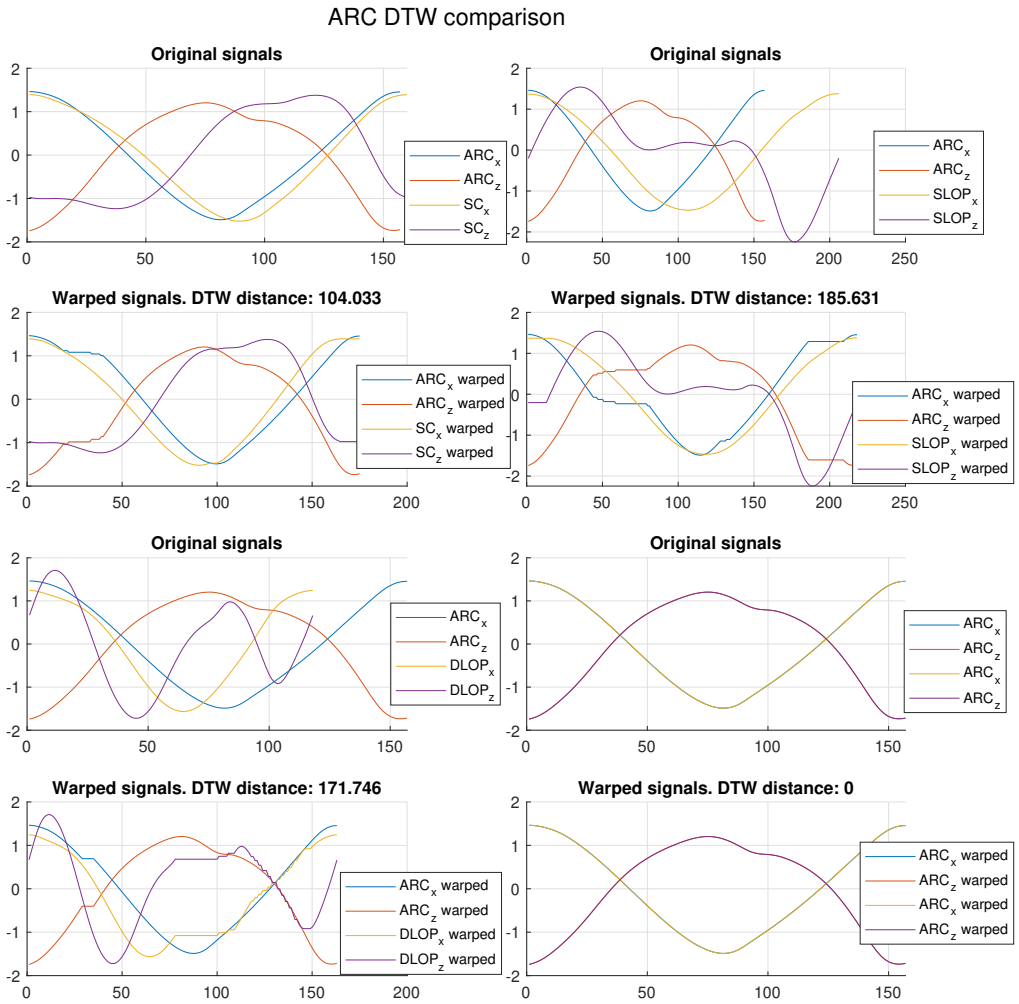


Figure A.6: Comparison of propulsion patterns using DTW analysis. Here, all shapes are compared to the ARC pattern.

A.2 Cycle Features

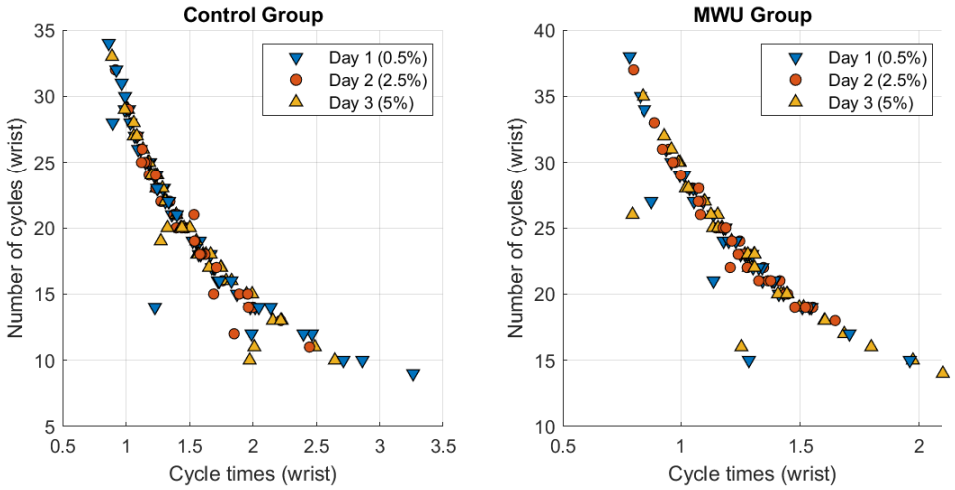
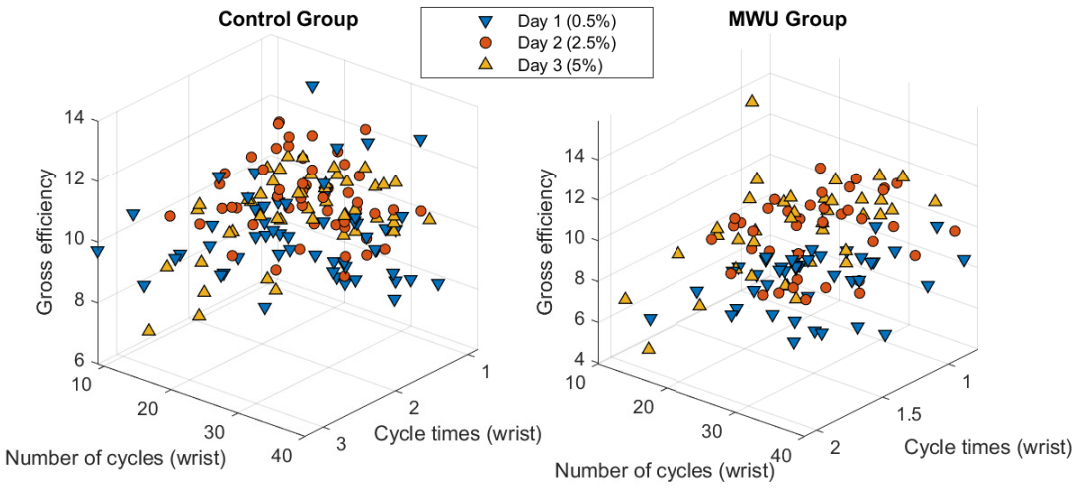
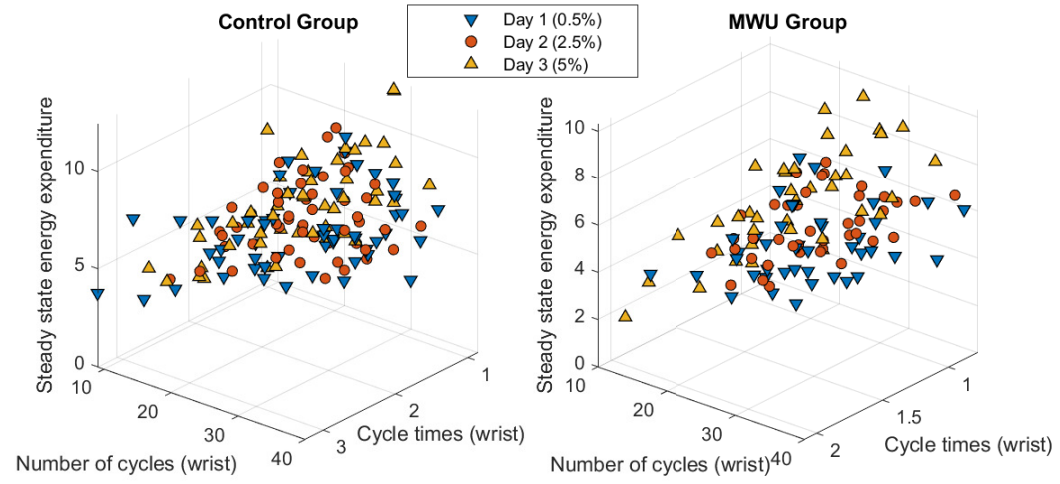


Figure A.7: Cycle features from the wrist marker. Labeled by experiment day.

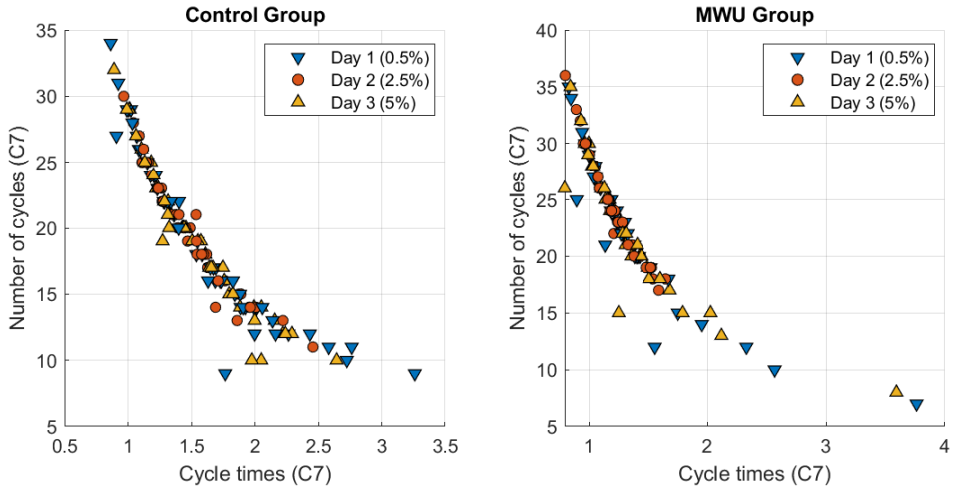


(a) Gross efficiency.

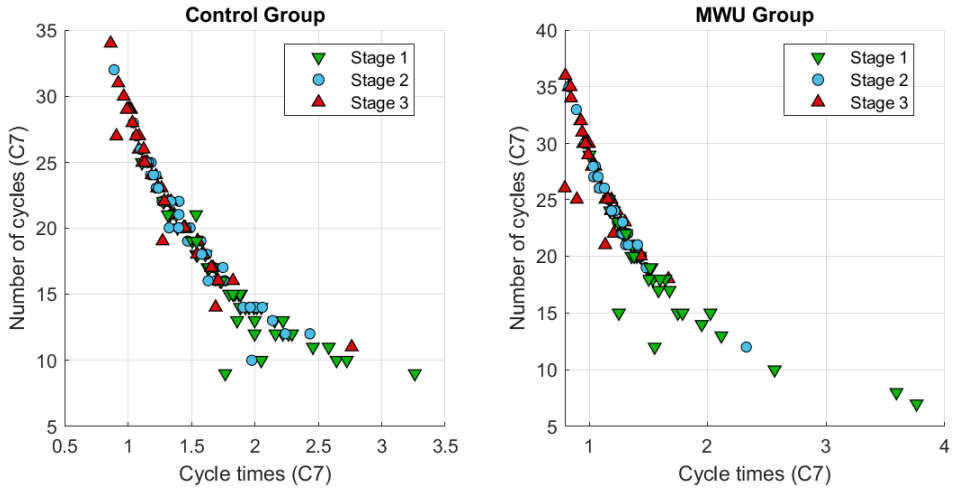


(b) Energy expenditure.

Figure A.8: Cycle features from the wrist marker plotted against gross efficiency and energy expenditure. Labeled by experiment day.

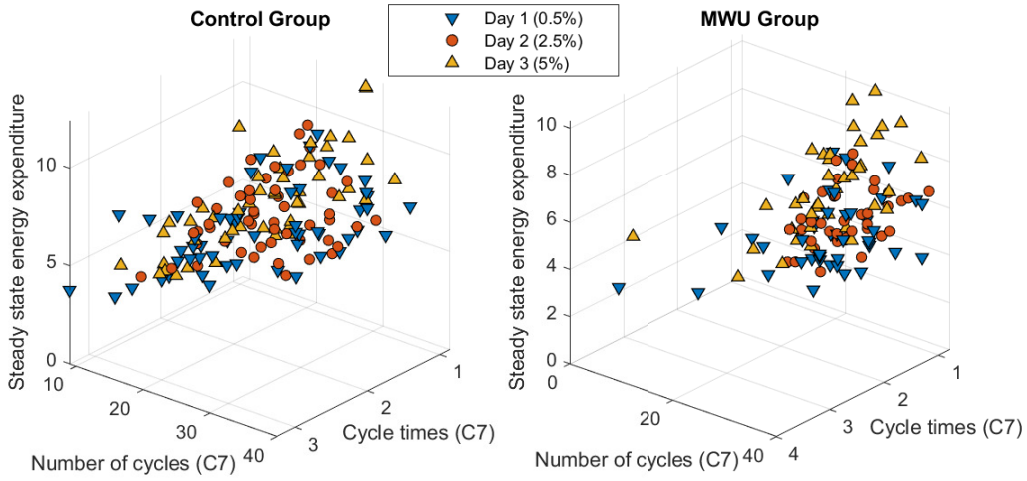


(a) Labeled by experiment day.

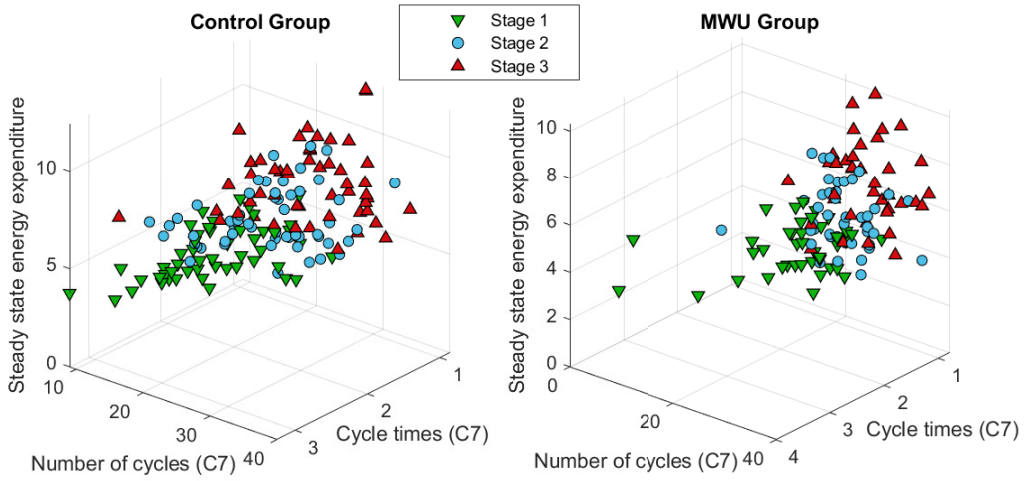


(b) Labeled by stage.

Figure A.9: Cycle features from the C7 marker.

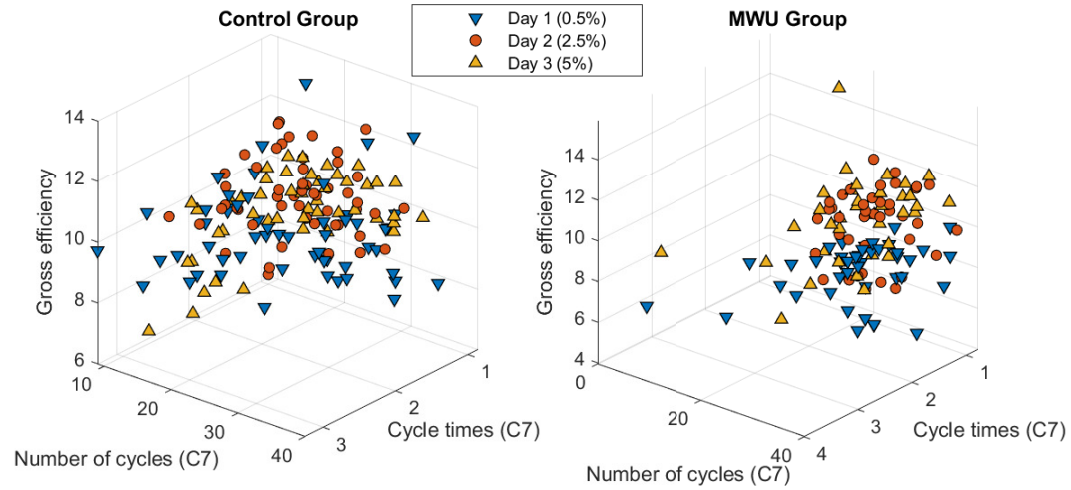


(a) Labeled by experiment day.

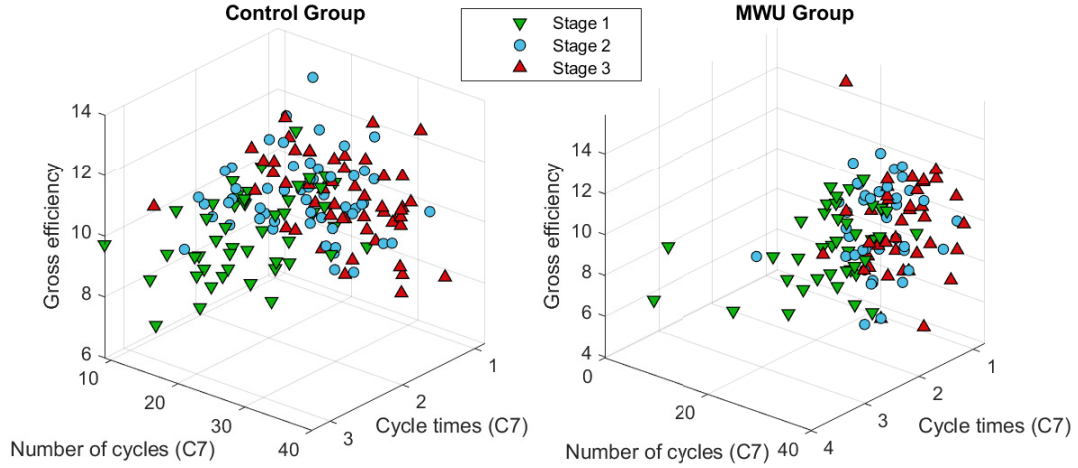


(b) Labeled by stage.

Figure A.10: Cycle features from the C7 marker plotted against energy expenditure.



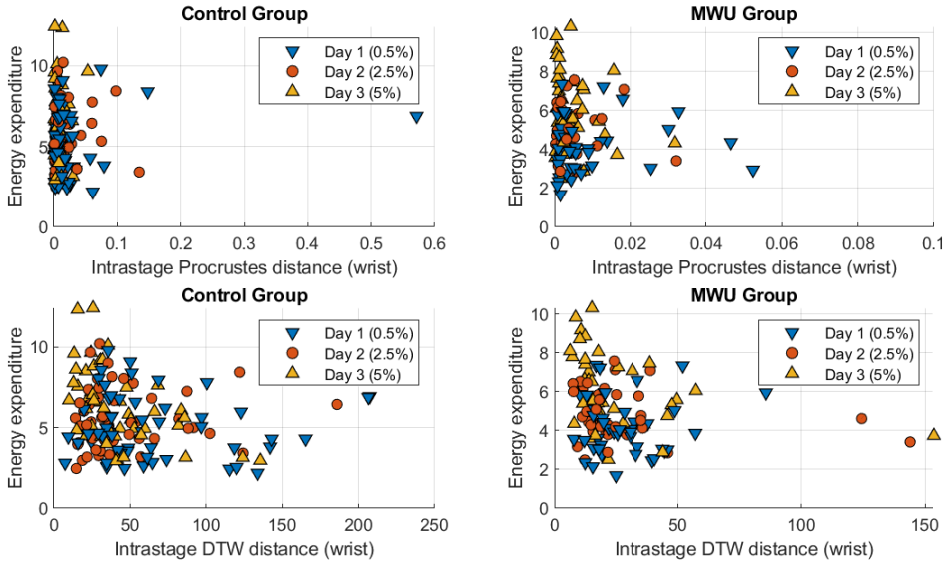
(a) Labeled by experiment day.



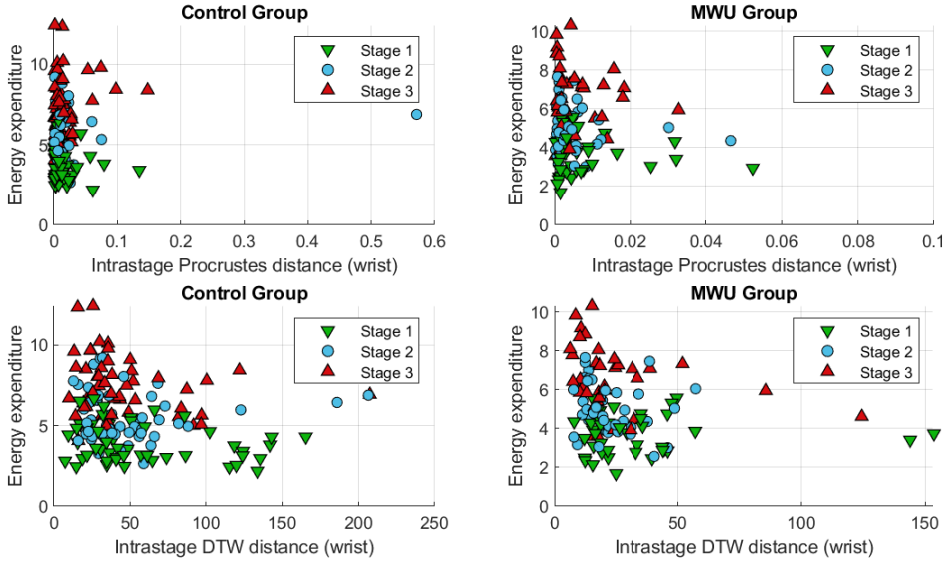
(b) Labeled by stage.

Figure A.11: Cycle features from the C7 marker plotted against gross efficiency.

A.3 Intrastage Variability

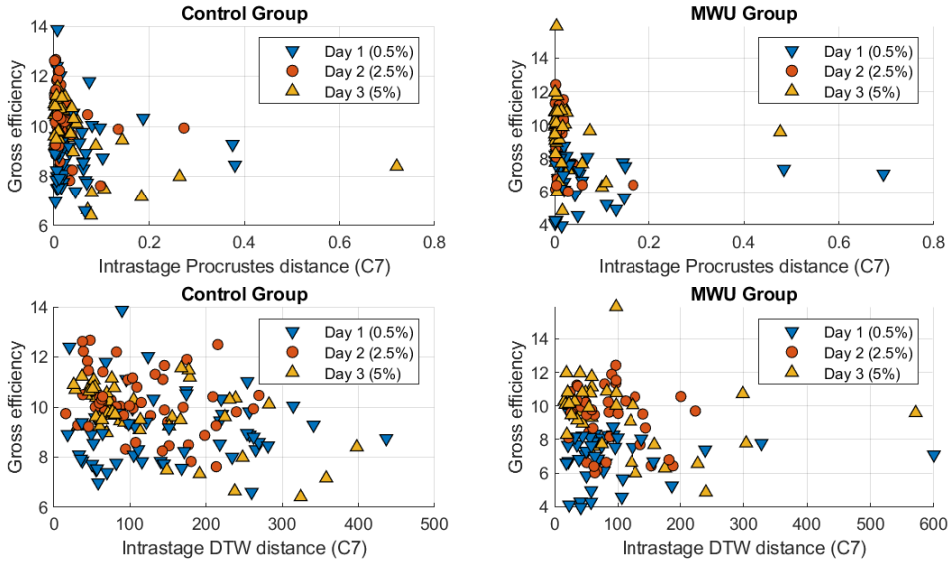


(a) Labeled by experiment day.

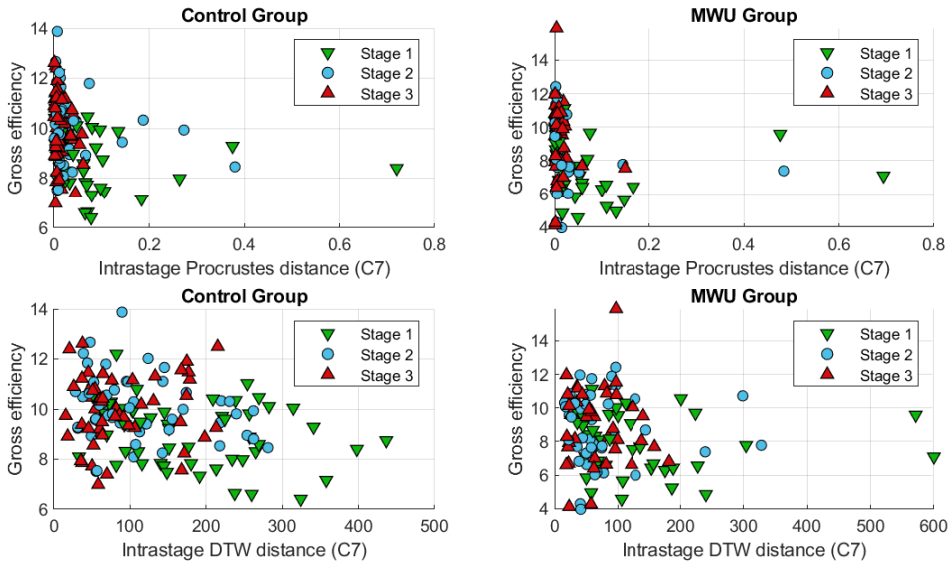


(b) Labeled by stage.

Figure A.12: Intrastage variability metrics from the wrist marker plotted against EE.



(a) Labeled by experiment day.



(b) Labeled by stage.

Figure A.13: Intrastage variability metrics from the C7 marker plotted against GE.

A.4 Interstage Variability

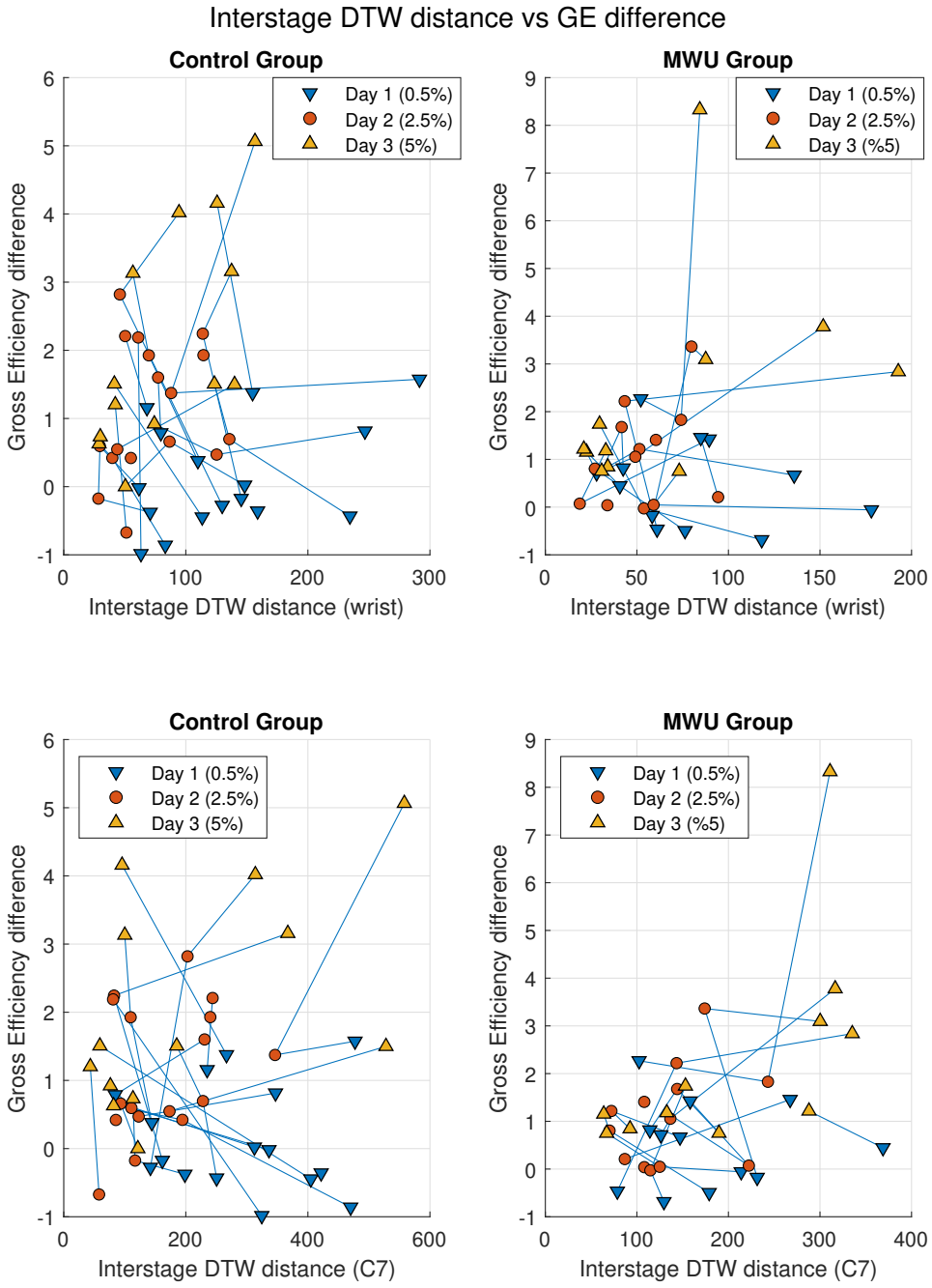


Figure A.14: Interstage DTW distances and differences in gross efficiency from stage 1 to stage 3, labeled by experiment day. Linked points represent a single participant.

A.5 Neural Network-Generated Trajectories

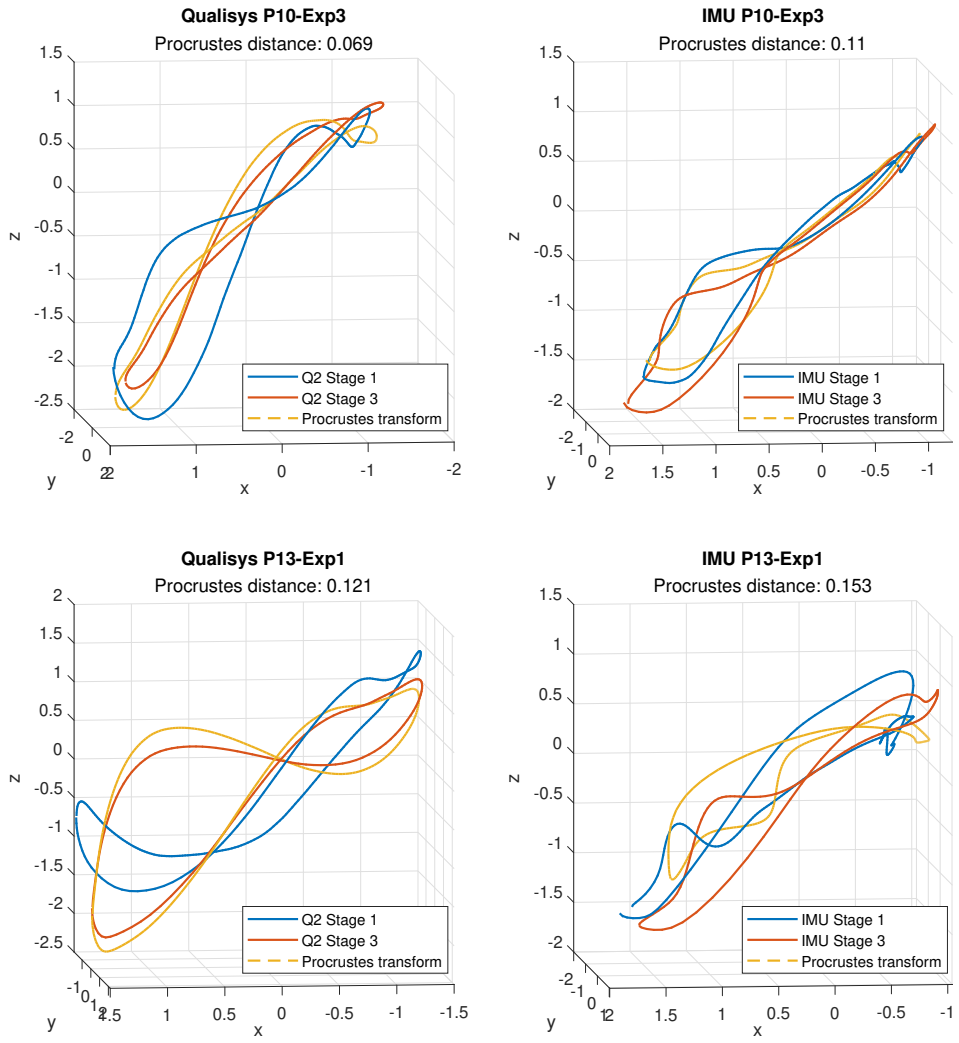


Figure A.15: Good replication of interstage variability with neural network-generated trajectories. Each row represents an individual example.

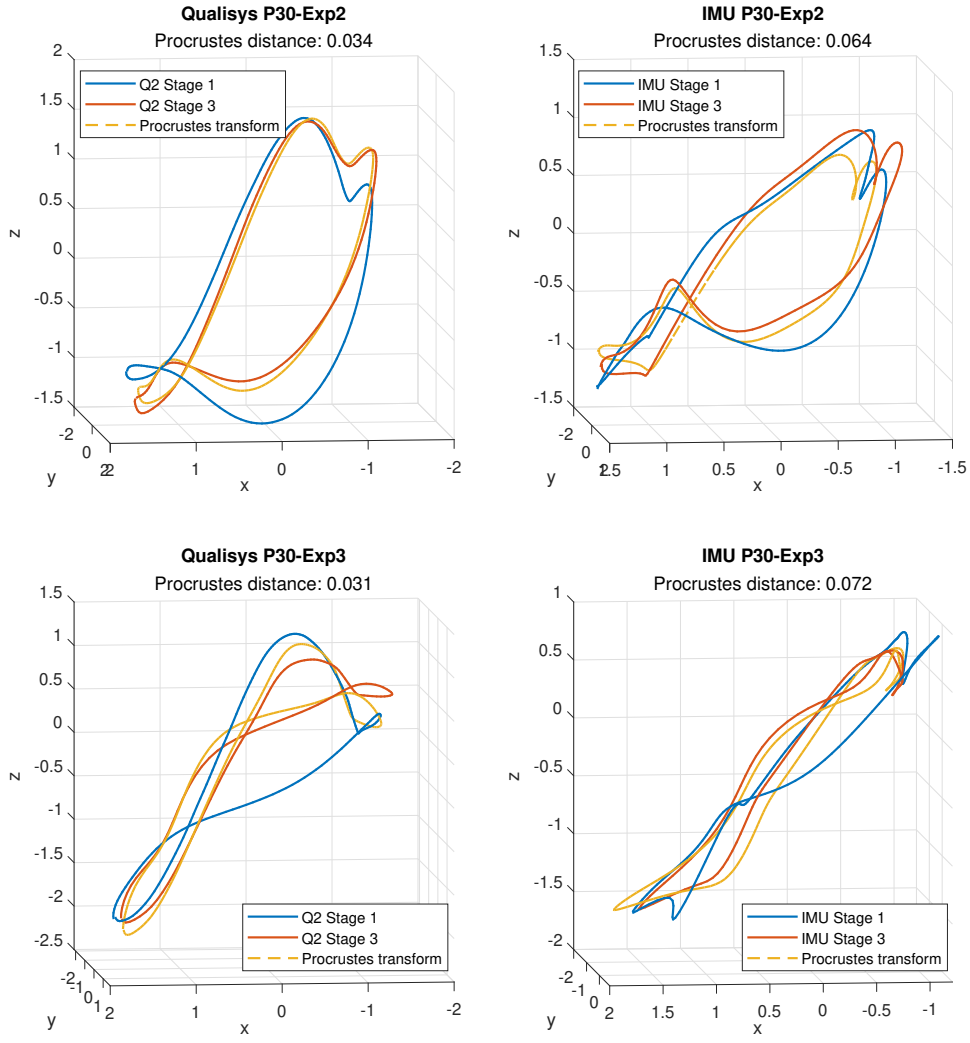


Figure A.16: Good replication of interstage variability with neural network-generated trajectories. Each row represents an individual example.

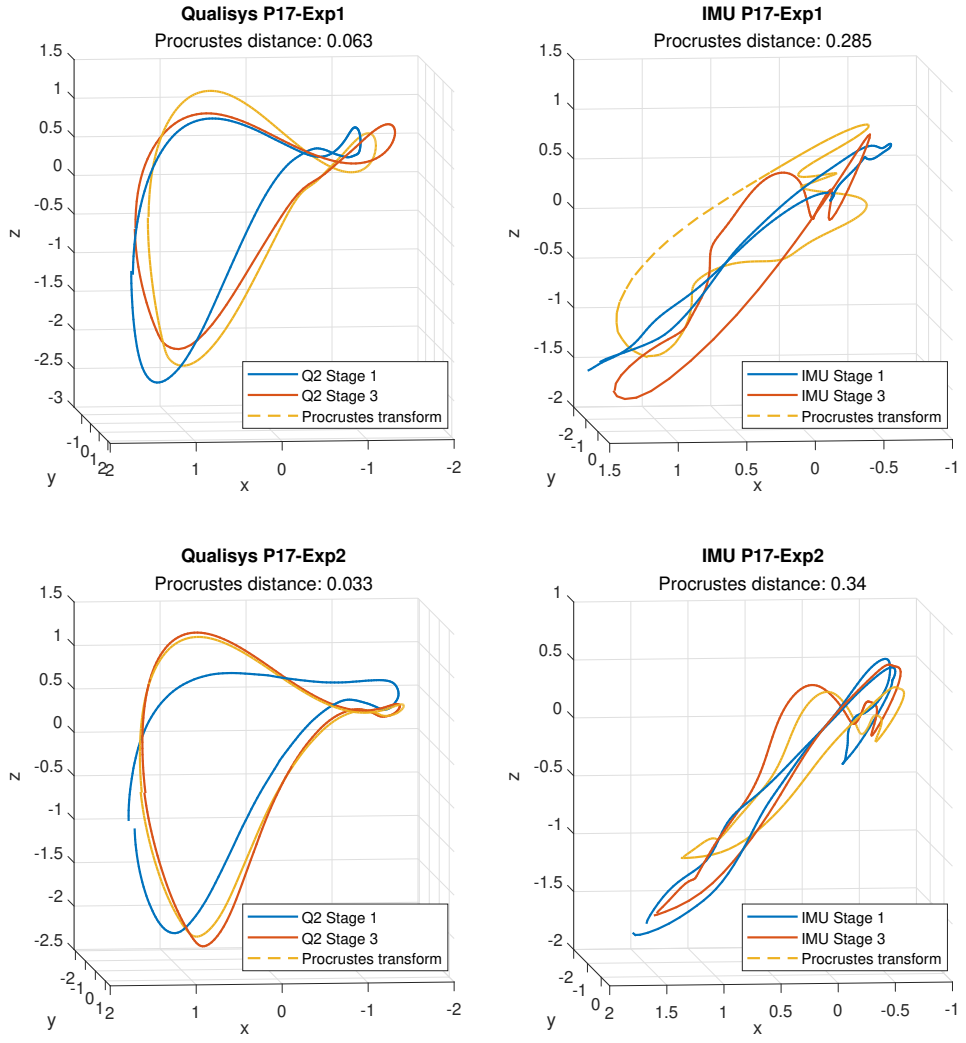


Figure A.17: Bad replication of interstage variability with neural network-generated trajectories. Each row represents an individual example.

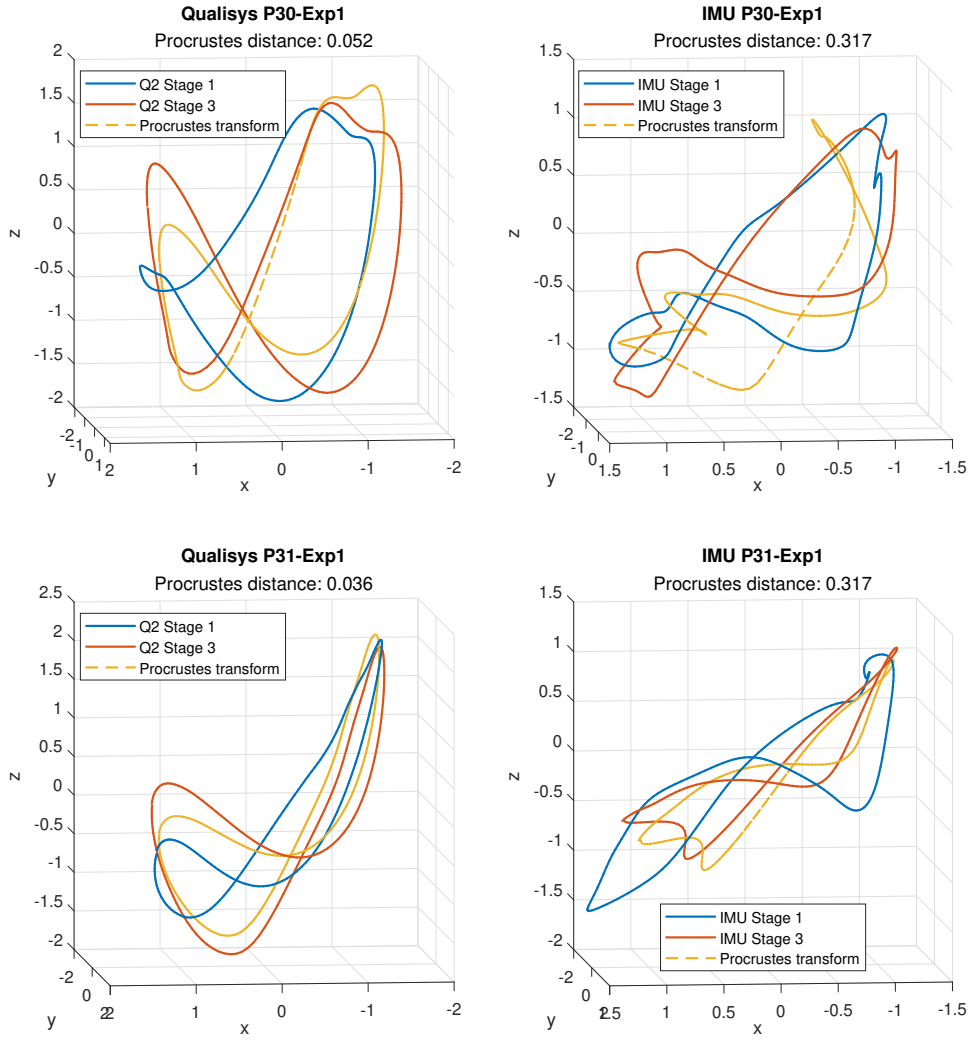


Figure A.18: Bad replication of interstage variability with neural network-generated trajectories. Each row represents an individual example.

NTNU

Norwegian University of
Science and Technology