



FIELD OF SCIENCE ENGINEERING AND TECHNOLOGY

SCIENTIFIC DISCIPLINE BIOMEDICAL ENGINEERING

DOCTORAL THESIS

Gait assessment by analyzing selected kinematic characteristics associated with the lower limbs movement in the sagittal plane

Author: Ludwin Molina Arias, MSc

Supervisor: Marek Iwaniec, PhD, DSc

Completed in: AGH University of Kraków, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, Department of Biocybernetics and Biomedical Engineering

Kraków, 2024



AGH

AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

DZIEDZINA NAUK INŻYNIERYJNO-TECHNICZNYCH

DYSCYPLINA INŻYNIERIA BIOMEDYCZNA

ROZPRAWA DOKTORSKA

Ocena chodu poprzez analizę wybranych charakterystyk
kinematycznych ruchu kończyny dolnej

Autor: mgr inż. Ludwin Molina Arias

Promotor rozprawy: dr hab. inż. Marek Iwaniec, Prof. AGH

Praca wykonana: Akademia Górniczo-Hutnicza, Wydział Elektrotechniki, Automatyki,
Informatyki i Inżynierii Biomedycznej, Katedra Biocybernetyki i Inżynierii Biomedycznej

Kraków, 2024

I would like to express my deepest gratitude to my supervisor, Marek Iwaniec, PhD, DSc, for his invaluable guidance and support during the preparation of this dissertation. His insightful advice, unwavering kindness, and the positive working atmosphere he has cultivated have been instrumental in the completion of this work. I am also deeply grateful to Joanna Iwaniec, PhD, DSc, for her continued encouragement, her constant availability to review my work, and for thoughtful discussions that have greatly enriched my academic journey. My sincere thanks are also extended to Prof. Piotr Augustyniak, PhD, DSc, for granting me the opportunity to conduct my research in the laboratories of the Department of Biocybernetics and Biomedical Engineering. Lastly, I wish to acknowledge the unconditional support of my partner, Laura Fernandez, and my parents, whose presence and encouragement have been a constant source of strength.

Abstract

Gait analysis, in a medical context, contributes to the identification of structural and functional limitations, as well as to the development of treatment plans and the evaluation of their efficacy in patients. Gait disorders, which can be the result of episodic events, such as unexpected falls, or chronic conditions, such as neurological dysfunctions, have a significant impact on people's lifestyles. They contribute to fall-related injuries, which are of great concern because of human suffering and associated economic costs. To address this problem, continuous advances are being made in gait analysis and pattern discrimination techniques.

Traditional quantitative gait assessment methodologies consist of examining spatial and/or temporal gait characteristics. These characteristics are either measured directly or derived from acquired data and compared to reference values of control subjects. Gait data can be scalar measurements, such as mean gait velocity, or time-dependent measurements, such as kinematic data and muscle activation signals. Since relying on scalar data alone is often insufficient to characterize a subject's gait, and relying on muscle dynamics analysis requires considerable effort, time, and specialized equipment, there is growing interest in gait analysis approaches that focus on kinematic data.

Kinematic gait analysis focuses primarily on examining the movements of the body's joints. Although the motion patterns of the major joints during gait are well documented and their efficacy in gait analysis is established, other promising kinematic measures have received less attention. One such measure is the Instantaneous Center of Rotation (ICR) motion of body segments, which may offer valuable information beyond traditional gait studies.

The aim of this research is to evaluate the potential of ICR trajectories of lower extremity segments in the sagittal plane as a novel parameter to characterize gait and diagnose gait disorders. Within the framework of the work, an overview of the methods used for gait-related data acquisition is provided, as well as a review of current methodologies for gait assessment. To characterize normal gait, a simplified model of the human lower extremities is employed, considering each segment as a rigid body. As part of the study, gait trials were performed in a variety of settings, with participants walking on the ground at their preferred pace and on a treadmill at a controlled speed. Data acquisition was performed using sports cameras and a portable MoCap system. The ICR position of lower extremity segments throughout the gait cycle is estimated using a geometric approach, and the resulting ICR trajectories are used to characterize normal gait.

A novel methodology is proposed to discriminate abnormal gait patterns based on the studied ICR trajectories. This methodology consists of determining the ICR position of lower extremity segments over time and representing each gait trial as a multidimensional time series. Pairwise comparisons of these time series using dissimilarity measures, together with a self-organization method, are used to generate a three-

dimensional network. In this network, the nodes represent walking trials and the spatial distances between them reflect their dissimilarities. Finally, cluster analysis is used to automatically identify groups of nodes, and thus subjects, with similar gait patterns, and to discriminate outlier nodes, possibly representing subjects with gait abnormalities.

To evaluate the sensitivity of the proposed kinematic parameter, difficulties were intentionally introduced to individuals during the experiments, revealing a correlation between increasing gait difficulty and a parameter derived from ICR trajectories called global gait feature. In addition, the proposed methodology was tested on a database of patients with Parkinson's disease, yielding promising results.

Streszczenie

Analiza chodu w kontekście medycznym przyczynia się do identyfikacji ograniczeń strukturalnych i funkcjonalnych, a także do opracowania planów leczenia i oceny ich skuteczności u pacjentów. Zaburzenia chodu, które mogą być wynikiem epizodycznych zdarzeń, takich jak niespodziewane upadki, lub przewlekłych schorzeń, takich jak dysfunkcje neurologiczne, mają znaczący wpływ na styl życia ludzi. Przyczyniają się one do urazów związanych z upadkami, które są bardzo niepokojące ze względu na ludzkie cierpienie i związane z nimi koszty ekonomiczne. Aby zaradzić temu problemowi, dokonywane są ciągłe postępy w analizie chodu i technikach rozróżniania wzorców.

Tradycyjne metodologie ilościowej oceny chodu polegają na badaniu przestrzennych i/lub czasowych cech chodu. Cechy te są mierzone bezpośrednio lub na podstawie uzyskanych danych i porównywane z wartościami referencyjnymi osób kontrolnych. Dane dotyczące chodu mogą być pomiarami skalarnymi, takimi jak średnia prędkość chodu, lub pomiarami zależnymi od czasu, takimi jak dane kinematyczne i sygnały aktywacji mięśni. Ponieważ poleganie wyłącznie na danych skalarnych jest często niewystarczające do scharakteryzowania chodu pacjenta, a poleganie na analizie dynamiki mięśni wymaga znacznego wysiłku, czasu i specjalistycznego sprzętu, rośnie zainteresowanie metodami analizy chodu, które koncentrują się na danych kinematycznych.

Kinematyczna analiza chodu skupia się przede wszystkim na badaniu ruchów stawów ciała. Podczas gdy wzorce ruchu głównych stawów podczas chodu są dobrze udokumentowane, a ich skuteczność w analizie chodu jest ustalona, inne obiecujące miary kinematyczne otrzymały mniej uwagi. Jedną z takich miar jest natychmiastowy ruch środka obrotu (ICR) segmentów ciała, który może oferować cenne informacje wykraczające poza tradycyjne badania chodu.

Celem niniejszego badania jest ocena potencjału trajektorii ICR segmentów kończyn dolnych w płaszczyźnie strzałkowej jako nowego parametru charakteryzującego chód, przydatnego do diagnozowania zaburzeń chodu. W ramach pracy przedstawiono przegląd metod stosowanych do pozyskiwania danych związanych z chodem, a także przegląd aktualnych metodologii oceny chodu. Aby scharakteryzować normalny chód, zastosowano uproszczony model ludzkich kończyn dolnych, traktując każdy segment jako ciało sztywne. W ramach badania przeprowadzono próby chodu w różnych warunkach, z uczestnikami chodzącymi po ziemi w preferowanym tempie i na bieżni z kontrolowaną prędkością. Gromadzenie danych odbywało się za pomocą kamer sportowych i przenośnego systemu MoCap. Pozycja ICR segmentów kończyn dolnych w całym cyklu chodu została oszacowana przy użyciu podejścia geometrycznego, a wynikowe trajektorie ICR zostały wykorzystane do scharakteryzowania normalnego chodu.

Zaproponowano nowatorską metodologię rozróżniania nieprawidłowych wzorców chodu na podstawie badanych trajektorii ICR. Metodologia ta polega na określeniu pozycji ICR segmentów kończyn dolnych w czasie i reprezentowaniu każdej próby chodu jako wielowymiarowego szeregu czasowego. Porównania parami tych szeregów czasowych przy użyciu miar niepodobieństwa, wraz z metodą samoorganizacji, są wykorzystywane do generowania trójwymiarowej sieci. W tej sieci węzły reprezentują próby chodzenia, a odległości przestrzenne między nimi odzwierciedlają ich odmienność. Wreszcie, analiza skupień jest wykorzystywana do automatycznej identyfikacji grup węzłów, a tym samym podmiotów, o podobnych wzorcach chodu oraz do rozróżniania węzłów odstających, prawdopodobnie reprezentujących podmioty z nieprawidłowościami chodu.

Aby ocenić wrażliwość proponowanego parametru kinematycznego, podczas eksperymentów celowo wprowadzono trudności dla poszczególnych osób, ujawniając korelację między rosnącą trudnością chodu a parametrem pochodzącym z trajektorii ICR zwanym globalną cechą chodu. Ponadto proponowana metodologia została przetestowana na bazie danych pacjentów z chorobą Parkinsona, dając obiecujące wyniki.

Contents

List of Figures	xv
List of Tables	xxi
List of Acronyms	xxiii
1 Introduction	1
1.1 Background and motivation	1
1.2 Research aim, scope, and hypotheses	3
1.3 Thesis outline	3
2 Biomechanics of normal gait	5
2.1 The gait cycle	5
2.2 Spatiotemporal gait parameters	8
2.3 Gait kinematics	10
2.3.1 Pelvis kinematics	10
2.3.2 Hip joint kinematics	11
2.3.3 Knee joint kinematics	12
2.3.4 Ankle joint kinematics	12
2.4 Gait kinetics	13
2.4.1 Ground Reaction Forces	13
2.4.2 Hip joint kinetics	15
2.4.3 Knee joint kinetics	16
2.4.4 Ankle joint kinetics	17
3 Overview on kinematic gait analysis	19
3.1 Introduction	19
3.2 Methods for kinematic data acquisition	22
3.2.1 Vision-based methods	22
3.2.2 Wearable devices	24
3.2.3 Commercial Motion Capture systems	27
3.3 Review on kinematic-based gait assessment	29
4 Characterization of human gait by means of ICR trajectories	33

4.1	Introduction	33
4.1.1	Principle of planar kinematics	34
4.2	Computation of ICR trajectories of the lower limb segments	38
4.3	Gait-related kinematic data acquisition	39
4.3.1	Subjects	39
4.3.2	Instrumentation	39
4.3.3	Experimental protocol	43
4.4	Results and discussion	45
4.4.1	Analysis of ICR trajectories in time domain	50
4.4.2	Frequency content of ICR trajectories	64
4.4.3	Analysis of ICR trajectories in state-space	69
5	A novel methodology to assess human gait	81
5.1	Introduction	81
5.2	General principle	82
5.3	Gait-related kinematic data acquisition and ICR computation	83
5.4	Pairwise comparison of gait trials	84
5.4.1	Data preprocessing	84
5.4.2	Pairwise comparison of time series using similarity measures	86
5.4.3	Pairwise comparison of time series using distance-based metrics	87
5.4.4	Ensemble of the dissimilarity matrix	90
5.5	Self-organization method for data embedding in 3D space	91
5.6	Clustering of gait-related data	93
5.7	Results and discussion	95
5.7.1	Sensitivity analysis	95
5.7.2	Exploring the entire dataset	108
6	Applications of ICR-LLS to discriminate individuals with Parkinson Disease	115
6.1	Introduction	115
6.2	Selected kinematic gait data sets	116
6.2.1	Gait data set on patients with PD	116
6.2.2	Gait data set on healthy individuals	117
6.3	Preparation of dataset for analysis	119
6.4	Results and discussion	121
7	Conclusions and final remarks	137
	Bibliography	141
A	Experimental Protocol	153
A.1	Sensor placement in MVN Awinda System	155

A.1.1	Head sensor	155
A.1.2	Shoulder sensors	155
A.1.3	Upper arm sensors	156
A.1.4	Forearm sensors:	156
A.1.5	Hand sensors:	156
A.1.6	Pelvis sensor:	157
A.1.7	Sternum sensor:	157
A.1.8	Upper leg sensors:	157
A.1.9	Lower leg sensors:	158
A.1.10	Foot sensors:	158
A.2	Marker placement for vision-based method	159
A.2.1	Hip joint marker	159
A.2.2	Knee joint marker	159
A.2.3	Ankle: Lateral Maleolus marker	159
A.2.4	Foot metatarsal head marker	159
B	Application of ICR-LLS methodology in the diagnosis of PD	161
B.1	Computed dissimilarity matrices	161
B.2	Cluster analysis	162

List of Figures

2.1	Division of the GC.	6
2.2	Position of right leg (black) at (a) Initial contact and (b) Loading response.	7
2.3	Position of right leg at Mid stance and Terminal Stance.	7
2.4	Position of right leg at Pre-swing and Early swing.	8
2.5	Position of right leg (black) at (a) Mid swing and (b) Terminal swing.	8
2.6	Spatial gait parameters.	9
2.7	Angular displacement of the pelvis joint in the sagittal and frontal plane throughout the GC.	11
2.8	Angular displacement of the hip joint in the sagittal and frontal plane throughout the GC.	11
2.9	Angular displacement of the knee joint in the sagittal and frontal plane throughout the GC.	12
2.10	Angular displacement of the ankle joint in the frontal plane throughout the GC.	13
2.11	Ground reaction forces (GRF) throughout the GC.	14
2.12	Internal torque and power of the hip joint in the sagittal and frontal plane throughout the GC.	15
2.13	Internal torque and power of the knee joint in the sagittal and frontal plane throughout the GC.	16
2.14	Internal torque and power of the ankle joint in the sagittal and frontal plane throughout the GC.	17
3.1	JAKC OGA form: ankle, foot, and toe deviations.	19
3.2	Instrumentation used for gait-related data acquisition in a typical gait laboratory.	21
3.3	Use of vision-based methods in gait analysis.	22
3.4	Use of optical-based methods in gait analysis.	23
3.5	Lumped model of an accelerometer.	24
3.6	Lumped model of a gyroscope.	25
3.7	Use of flexible goniometers for measuring knee joint flexion/extension.	26
3.8	Setup of a contemporary gait analysis laboratory.	27
3.9	Inertial MoCap suit developed by Movella.	28
3.10	Main stages of instrumented gait analysis in the diagnosis of gait anomalies.	29
3.11	Gait assessment pipeline.	30
3.12	ICR trajectory of intervertebral segments during weight-bearing tasks.	31
4.1	Rigid body undergoing pure translation.	34
4.2	Rigid body undergoing pure rotation around point O	34
4.3	Rigid body undergoing general planar motion consisting of a translation from O_1 to O_2 and a rotation about O	35

4.4	Determination of the position of a particle belonging to a rigid body with respect to an inertial reference system.	35
4.5	Determination of the ICR position of a rigid body undergoing general planar motion.	37
4.6	Approaches used for the calculation of the position of a rigid body ICR.	38
4.7	Use of Kinovea for tracking lower limb joint angles.	40
4.8	Technical parameters of sport camera GoPro 10.	41
4.9	Technical parameters of sport camera GoPro 8.	41
4.10	MVN Awinda MoCap system.	42
4.11	Setup for the acquisition of gait-related kinematic data.	43
4.12	Marker used for kinematic gait-related data acquisition using Kinovea.	44
4.13	Placement of the trackers and markers.	45
4.14	Hip marker trajectory on the vertical axis.	46
4.15	Trajectories of the ankle, knee and hip markers in the horizontal and vertical axes throughout the GC.	47
4.16	Angular displacement of the knee joint throughout the GC.	48
4.17	Velocity of the ankle, knee and hip markers in the horizontal and vertical axes throughout the GC.	49
4.18	ICR position of the shank and thigh around (a) 5%, (b) 35%, (c) 65% and (d) 95% of the GC	50
4.19	Shank ICR trajectory during treadmill gait trials.	52
4.20	Thigh ICR trajectory during different gait trials.	53
4.21	Comparison of shank ICR trajectories in Cartesian coordinates during treadmill and over-ground gait trials using different data acquisition techniques.	54
4.22	Comparison of thigh ICR trajectories in Cartesian coordinates during treadmill and over-ground gait trials using different data acquisition techniques.	54
4.23	Comparison of shank ICR trajectories in polar coordinates during treadmill and overground gait trials using different data acquisition techniques.	56
4.24	Comparison of thigh ICR trajectories in polar coordinates during treadmill and overground gait trials using different data acquisition techniques.	56
4.25	Shank ICR trajectories in Cartesian coordinates of each subject during treadmill gait trials. .	58
4.26	Thigh ICR trajectories in Cartesian coordinates of each subject during treadmill gait trials. .	58
4.27	Shank ICR trajectories in Cartesian coordinates of each subject during overground gait trials.	59
4.28	Thigh ICR trajectories in Cartesian coordinates of each subject during overground gait trials.	59
4.29	Segmented gait intervals obtained from data collected during treadmill gait trials using the MVN Awinda system.	60
4.30	Segmented gait intervals obtained from data collected during overground gait trials using the MVN Awinda system.	61
4.31	Segmented gait intervals obtained from data collected during treadmill gait trials using Kinovea.	62
4.32	Planar ICR trajectory of shank and thigh during different gait trials.	63

4.33	Principle of Short-Time Fourier Transform.	65
4.34	Frequency content of ICR trajectories during treadmill gait trials collected using the MVN Awinda system.	66
4.35	Frequency content of ICR trajectories during overground gait trials collected using the MVN Awinda system	66
4.36	Frequency content of ICR trajectories during treadmill gait trials collected using Kinovea. . .	67
4.37	Spectrogram of ICR trajectories during treadmill gait trials collected using the MVN Awinda system.	67
4.38	Spectrogram of ICR trajectories during overground gait trials collected using the MVN Awinda system.	68
4.39	Spectrogram of ICR trajectories during treadmill gait trials collected using Kinovea.	68
4.40	Phase portrait of a simple pendulum.	70
4.41	Time delay embedding procedure.	71
4.42	Characteristic typology of RPs.	74
4.43	RPs obtained from shank and thigh ICR trajectories.	76
4.44	RPs obtained from combined shank and thigh ICR trajectories.	76
5.1	Flow chart of the ICR-LLS methodology showing the sequential steps involved.	82
5.2	Heel added to the sole of the subjects during gait trials to simulate different levels of gait abnormalities.	83
5.3	Calculation of the warping path by using dynamic programming.	89
5.4	Force-directed graph layout drawing technique.	92
5.5	DBSCAN point types.	93
5.6	Shank ICR trajectories during treadmill gait trials for different levels of gait difficulty. . . .	96
5.7	Shank ICR trajectories during overground gait trials for different levels of gait difficulty. . .	96
5.8	Dissimilarity matrices calculated from pairwise comparisons of shank ICR trajectories of the same subject during treadmill gait trials.	98
5.9	Dissimilarity matrices calculated from pairwise comparisons of shank ICR trajectories of the same subject during overground gait trials.	99
5.10	Networks generated from data collected during treadmill gait trials of the same subject at different difficulty levels.	100
5.11	Networks generated from data collected during overground gait trials of the same subject at different difficulty levels.	101
5.12	Extracted gait features from data collected during treadmill gait trials at different difficulty levels.	102
5.13	Extracted gait features from data collected during overground gait trials at different difficulty levels.	103
5.14	Average gait features extracted from data collected during treadmill and overground gait trials.	104
5.15	Single gait feature of the different level of gait difficulties extracted from the data collected during treadmill and overground gait trials.	105

5.16	Recalculated dissimilarity metrics based on the layouts of the network generated from the data acquired during treadmill gait trials.	106
5.17	Recalculated dissimilarity metrics based on the layouts of the network generated from the data acquired during overground gait trials.	107
5.18	Dissimilarity matrices calculated from pairwise comparisons of shank ICR trajectories of all subjects during treadmill gait trials.	108
5.19	Dissimilarity matrices calculated from pairwise comparisons of shank ICR trajectories of all subjects during overground gait trials.	109
5.20	Node arrangement generated from data collected during treadmill gait trials under different difficulty levels and using the Förstner distance for comparisons of multivariate time series. .	110
5.21	Node arrangement generated from data collected during overground gait trials under different difficulty levels and using the Förstner distance for comparisons of multivariate time series.	111
5.22	Node arrangement generated from data collected during treadmill gait trials under different difficulty levels and using the Förstner distance for comparisons of multivariate time series. .	112
5.23	Node arrangement generated from data collected during overground gait trials under different difficulty levels and using the Förstner distance for comparisons of multivariate time series.	113
6.1	Reference frames of the anatomical markers on the PD patients.	117
6.2	Reference frames of the anatomical markers on the healthy individuals.	118
6.3	Processing of the kinematic data contained in the selected data sets using Mokka.	120
6.4	Thigh ICR trajectories of patients with PD during overground gait trials in Cartesian coordinates.	121
6.5	Shank ICR trajectories of patients with PD during overground gait trials in Cartesian coordinates.	122
6.6	Thigh ICR trajectories of patients with PD during overground gait trials in polar coordinates.	122
6.7	Shank ICR trajectories of patients with PD during overground gait trials in polar coordinates.	123
6.8	Comparison of shank ICR trajectories of subjects with and without PD during overground gait trials.	124
6.9	Dissimilarity matrix calculated from pairwise comparisons of the shank ICR trajectories of subjects with and without PD using Förstner distance.	125
6.10	Network layout estimated from the Förstner distance-based dissimilarity matrix.	126
6.11	Grid constructed to search for optimal tuning of the clustering algorithm parameters.	127
6.12	Clusters identified from the Förstner distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.04$ and $MinPts = 5$	128
6.13	Silhouette values for element clustering derived from the Förstner distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 0.87$ and $MinPts = 3$	129
6.14	Clusters identified from the Förstner distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 0.99$ and $MinPts = 4$	131

6.15	Co-association matrix computed from multiple simulations with varying clustering algorithm parameters.	132
6.16	Clusters identified from the Förstner distance-based network layout using the DBSCAN algorithm and clustering ensemble approach.	132
B.1	Dissimilarity matrices calculated from pairwise comparisons of the shank ICR trajectories of subjects with and without PD.	161
B.2	Clusters identified from the Euclidean distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.79$ and $MinPts = 7$	162
B.3	Clusters identified from the Euclidean distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.61$ and $MinPts = 5$	162
B.4	Clusters identified from the Mean dissimilarity-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.35$ and $MinPts = 5$	163
B.5	Clusters identified from the Mean dissimilarity-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.52$ and $MinPts = 3$	163
B.6	Clusters identified from the RMS dissimilarity-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.47$ and $MinPts = 6$	164
B.7	Clusters identified from the RMS dissimilarity-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.29$ and $MinPts = 3$	164
B.8	Clusters identified from the Peak dissimilarity-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.61$ and $MinPts = 6$	165
B.9	Clusters identified from the Peak dissimilarity-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.42$ and $MinPts = 4$	165
B.10	Clusters identified from the Manhattan distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.59$ and $MinPts = 6$	166
B.11	Clusters identified from the Manhattan distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.47$ and $MinPts = 6$	166
B.12	Clusters identified from the DTW-based network layout using the DBSCAN algorithm with parameters $\epsilon = 0.92$ and $MinPts = 5$	167
B.13	Clusters identified from the DTW-based network layout using the DBSCAN algorithm with parameters $\epsilon = 0.91$ and $MinPts = 4$	167
B.14	Clusters identified from the Cross-Correlation-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.22$ and $MinPts = 3$	168
B.15	Clusters identified from the Cross-Correlation-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.07$ and $MinPts = 3$	168
B.16	Clusters identified from the Cross-Correlation + DTW-based network layout using the DBSCAN algorithm with parameters $\epsilon = 0.89$ and $MinPts = 4$	169
B.17	Clusters identified from the Cross-Correlation + DTW-based network layout using the DBSCAN algorithm with parameters $\epsilon = 0.85$ and $MinPts = 3$	169

B.18 Clusters identified from the Förstner distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.04$ and $MinPts = 5$	170
B.19 Clusters identified from the Förstner distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 0.99$ and $MinPts = 4$	170

List of Tables

4.1	Demographic information regarding the participants involved in the gait trials.	40
4.2	Estimated embedding parameters for the dynamic systems of the shank and thigh.	73
4.3	Estimated embedding parameters for the dynamic shank-thigh system.	73
4.4	Summary of the interpretations of the structures found in the RPs.	75
4.5	RQA measures describing the structures in the RPs presented in Fig. 4.43	78
4.6	RQA measures describing the structures in the RPs presented in Fig. 4.44	78
6.1	Information of the selected patients with PD and healthy individuals.	119
6.2	Summary of clustering performance evaluations across different dissimilarity measures and clustering algorithm settings.	130
6.3	Information on subjects with medical diagnosed PD and the corresponding diagnosis result- ing from using the ICR-LLS methodology.	133
6.4	Comparison of the ICR-LLS methodology results with previous studies on PD classification	134

List of Acronyms

AMI Average Mutual Information

ASIC Application-Specific Integrated Circuit

BMI Body Mass Index

CGAM Comprehensive Gait Asymmetry Metric

CoM Center of Mass

DBSCAN Density-Based Spatial Clustering of Applications with Noise

DET Determinism

DTW Dynamic Time Warping

EMG Electromyography

ENTR Shannon Entropy

FNN False Nearest Neighborhood

FT Fourier Transform

GC Gait Cycle

GRF Ground Reaction Forces

ICR Instantaneous Center of Rotation

IMU Inertial Measurement Unit

LAM Laminarity

L_{MAX} Longest Diagonal Line

MA Moving Average

MEMS Microelectromechanical systems

MoCap Motion Capture

NFOG-Q New Freezing of Gait Questionnaire

NMI Normalized Mutual Information

OGA Observational Gait Analysis

PSD Power Spectral Density

RMS Root Mean Square

PD Parkinson Disease

RP Recurrence Plot

RQA Recurrence Quantification Analysis

RR Recurrence Rate

SC Silhouette Coefficient

SDI StrapDown Integration

STFT Short Time Fourier Transform

TT Trapping Time

UPDRS Unified Parkinson's Disease Rating Scale

WA Wavelet Analysis

Chapter 1

Introduction

1.1 Background and motivation

Walking is the most used method of displacement throughout our lives. From an early age, we intuitively learn to move our body forward while maintaining balance. Once learned, walking is an activity that is performed intuitively and unconsciously, although it is still a complex process that involves the participation of several systems, including the nervous, muscular, and skeletal systems. The movement pattern that humans perform while walking is, in general, a characteristic that we all share with only slight differences between individuals. This specific movement pattern known as gait is studied for a variety of purposes, including the design of rehabilitation and gait assistance devices (Nuckols et al., 2020), recognition of people (Kumar and Babu, 2012), and the diagnosis of neurodegenerative and motor disorders (Salarian et al., 2004).

Gait patterns are influenced by several factors such as age, personality, mood, and sociocultural background (Pirker and Katzenschlager, 2017). For example, people living in urban areas tend to walk faster than those in rural regions (Ebersbach et al., 2000). Gait disorders are widespread and can be observed in more than half of non-bedbound patients in neurology inpatient services and more than 20% of geriatric patients, with a higher prevalence with age (Giladi, 2019). In fact, the frequency of gait and balance disorders increases significantly with age, ranging from approximately 10% in individuals aged 60 to 69 years, to more than 60% in those over 80 years (Mahlknecht et al., 2013). Understanding these patterns is vital for a comprehensive overview of gait analysis and its implications for various age groups and populations.

In medicine, gait analysis has proven to be a valuable tool in diagnosing and treating various medical conditions that affect the musculoskeletal system. Adverse health conditions, such as the risk of developing dementia and the risk of falling, which is the most common type of home accident (Rubenstein et al., 2004) and therefore a greatest threat to health and independence, are efficiently examined by gait analysis. Gait analysis plays a crucial role in the early diagnosis and monitoring of neurodegenerative diseases, with numerous studies highlighting the close connection between cognitive impairment and mobility alterations in these health conditions (Amboni et al., 2013). In patients with these diseases, the progressive degeneration of neuronal cells worsens with time, resulting in loss of neuromuscular control and disturbance of balance and gait (ataxia) or mental functioning (dementia). Accurate differentiation of gait disorders is crucial in

establishing the appropriate treatment plan, as misdiagnosis can have detrimental effects on patient quality of life.

To effectively assess gait and provide accurate diagnoses, researchers and clinicians rely primarily on kinematic, kinetic, and electromyography (EMG) data (Delisa, 1998). However, analyzing these data is not a simple task given their multidimensional, complex, and time-dependent nature. This complexity requires the application of sophisticated methodologies, such as data mining and machine learning, to extract meaningful insights and correlations from large data sets (Phinyomark et al., 2018). Furthermore, since the acquisition and analysis of kinetic and EMG data requires considerable effort, time, and specialized equipment, there is growing interest in gait analysis approaches that focus exclusively on kinematic data.

Gait assessment based on kinematic data has wide-ranging applications. In clinical settings, it has been shown to be capable of aiding in the diagnosis and monitoring of gait-related disorders, such as Parkinson's disease, cerebral palsy, or stroke (Kim et al., 2021). By comparing an individual's kinematic data with normative values, clinicians can identify deviations and design targeted interventions. Kinematic gait analysis is also essential in orthopedic rehabilitation, helping to assess the effectiveness of treatments and guide rehabilitation protocols. Furthermore, in sports performance analysis, kinematic data allows athletes and coaches to optimize technique, identify biomechanical inefficiencies, and reduce the risk of injuries (Aritan, 2015). The acquisition of kinematic data relies on the utilization of specialized equipment designed to accurately track body landmarks. These markers are used to measure and compute associated motion parameters, such as angular displacements of body segments, linear displacements of anatomical markers, and angular displacements of limb joints.

The analysis of gait kinematics commonly focuses on studying joint angular movements, which offer the advantage of being independent of the reference frame used for marker position measurements. Although the movement patterns of major body joints during gait are well documented, there are other parameters, such as the motion of the Instantaneous Center of Rotation (ICR) of body segments, that have received less attention. The ICR position of the lower limb segments, which varies throughout the gait cycle, particularly in the sagittal plane, can provide significant insight beyond conventional joint motion studies.

The primary focus of this research is to evaluate the potential of ICR trajectories of lower limb segments in the sagittal plane as a novel parameter to characterize gait and diagnose gait disorders, thus contributing to advances in gait analysis. The motivation for this research lies in the immense impact that gait analysis has on improving people's quality of life. By deepening the understanding of typical gait patterns and being able to detect gait deviations at an early stage, medical professionals can offer timely and targeted interventions to treat various medical conditions. With the immense advancement of data mining and the development of new machine learning techniques, the potential to extract relevant information from gait data is more promising than ever. Therefore, by using data mining to analyze new types of gait data, this work contributes to ongoing efforts to improve the diagnosis, treatment, and overall understanding of gait-related disorders. Ultimately, the goal of the work is to improve the well-being and mobility of people, especially those affected by neurodegenerative diseases, and to create a positive impact on their lives, both physically and emotionally.

1.2 Research aim, scope, and hypotheses

The objective of this study has been to evaluate to what extent the analysis of a commonly unused kinematic parameter in gait analysis, the instantaneous center of rotation of the segments of the lower extremities in the sagittal plane, can serve as a biomarker for diagnosing health conditions that alter normal gait patterns. To this end, a novel methodology for the discrimination of gait patterns has been proposed, which is based on pairwise comparisons of time series and cluster analysis and aims to discriminate subjects with gait disorders within groups containing control subjects. In relation to the research objective, the following research hypothesis was formulated:

It is possible to assess an individual's gait, distinguish gait disorders, and recognize health problems based on the analysis of the trajectories of the instantaneous center of rotation of lower limb segments in the sagittal plane during gait and pairwise pattern comparison using dissimilarity measures.

To substantiate the hypothesis, the study covers the following key aspects:

- ✓ The creation of a data set comprising gait-related kinematic data acquired from treadmill and over-ground gait trials performed on control subjects using a motion capture system and a vision-based method.
- ✓ Characterization of normal gait by means of the trajectories of the ICR of the lower limb segments on the sagittal plane throughout the gait cycle.
- ✓ The sensitivity analysis of the mean pattern of the ICR trajectories of the lower limb segments in the sagittal plane to gait perturbations induced in test subjects in a simulated manner.
- ✓ The proposal and testing of a methodology to distinguish gait patterns based on pairwise comparison and clustering of time series consisting of the ICR trajectories of lower limb segments in the sagittal plane.

1.3 Thesis outline

The thesis is made up of the following chapters:

1. **Introduction** – It provides an overview of the research problem to be solved and lays the foundation of the thesis by highlighting the importance of gait analysis in the diagnosis and treatment of various medical conditions. This chapter outlines the objective and scope of the investigation.
2. **Biomechanics of normal gait** – It describes the movement patterns present during human gait. The concept of the gait cycle is introduced and the phases that comprise it are described, detailing the major events that occur in each phase. The chapter also covers a review of the spatiotemporal parameters of gait and how to calculate them, including gait velocity, cadence, stride time, double support duration, stride length, stride length, stride width, and foot angle. In addition, it delves into the kinematics and

kinetics of the joints of the lower extremities, providing valuable information on normal gait patterns and the forces involved in gait movements.

3. **Overview on kinematic gait analysis** - Provides a review of the main methods used for the acquisition of kinematic data and their use in gait analysis for the diagnosis of health conditions involving gait disorders. This chapter presents the technology for studying gait used in gait clinics and laboratories and discusses specific examples of research using statistical or machine learning methods to extract valuable information from gait kinematic data.
4. **Characterization of human gait by means of the ICR trajectories of the lower limb segments in the sagittal plane** – It describes the procedure to determine the position of the instantaneous center of rotation of the lower limb segments in the sagittal plane. In addition, the protocol of the gait trials performed for the acquisition of kinematic data is described, as well as the demographic data of the test subjects. The time series pattern resulting from the determination of the ICR from the acquired gait data is analyzed in the time, frequency, and phase space domain, to characterize the average pattern present in normal gait.
5. **A novel methodology to assess human gait by matching ICR trajectories in the sagittal plane of lower limb segments** – It presents in detail the proposed methodology for discriminating gait patterns and covers each of its stages. Subsequently, the methodology is used to test the sensitivity of the ICR position of the lower limb segments in the sagittal plane when the subject's gait is affected. Finally, the performance of the methodology is evaluated to distinguish gait patterns within a data set of control subjects with different degrees of gait disturbance.
6. **Applications of ICR-LLS methodology to discriminate subjects with Parkinson disease** – In this chapter, the proposed methodology is tested using a dataset that contains kinematic data from subjects suffering from Parkinson's disease. The Parkinson data set is combined with the data set of acquired control subjects, and the methodology is used to distinguish one group from the other by distinguishing their gait patterns.
7. **Conclusions and final remarks** – Provides a summary of the entire dissertation.

Chapter 2

Biomechanics of normal gait

Gait is the result of a complex process that involves several systems of the human body, including the muscular, nervous, cardiovascular, and respiratory systems. These systems allow for the execution of body movements in a controlled manner while ensuring balance and stability. In the case of people with motor and neural disabilities, gait can be affected and can ultimately lead to falls, injuries, and loss of mobility. Eventually, these circumstances significantly reduce the quality of life of the affected person.

To understand pathological gait, it is necessary first to understand the biomechanics of normal gait, since this provides the standard against which the gait of a patient can be judged. Therefore, the purpose of this section is to assist the reader in understanding the mechanisms associated with normal gait.

2.1 The gait cycle

Gait comprises a series of rhythmic movements that are primarily executed by our limbs. The Gait Cycle (GC), also known as the stride, is defined as the time interval between two consecutive occurrences of one of the repetitive gait events. Although any event within the GC can be chosen indifferently, the instant at which the foot contacts the ground is usually used to define the GC. Therefore, the time interval between two foot strikes is the convention used in scientific and medical contexts to refer to GC ([Perry, 1992](#)).

In general, the GC is divided into two periods, known as the stance phase and the swing phase (see Fig. 2.1). This division is made according to whether there is contact between the foot and the ground. The stance phase refers to the period during which the foot is on the ground, while the swing phase refers to the period during which the foot moves forward in the air. Under normal conditions, the stance phase represents approximately 60% of the GC, and the swing phase represents the remaining 40%. However, the precise duration of a single GC varies with the person's walking speed.

The stance phase is subdivided into three intervals, corresponding to two periods of double stance and one period of single stance. The first double stance initiates the GC and is characterized by the contact of both feet with the ground. The single stance begins when the opposite foot is lifted to swing, and during this interval, the entire body weight rests on one limb. Finally, the terminal double stance begins with a single limb support and continues until the original supporting limb is lifted to swing ([Whittle, 2007](#)).

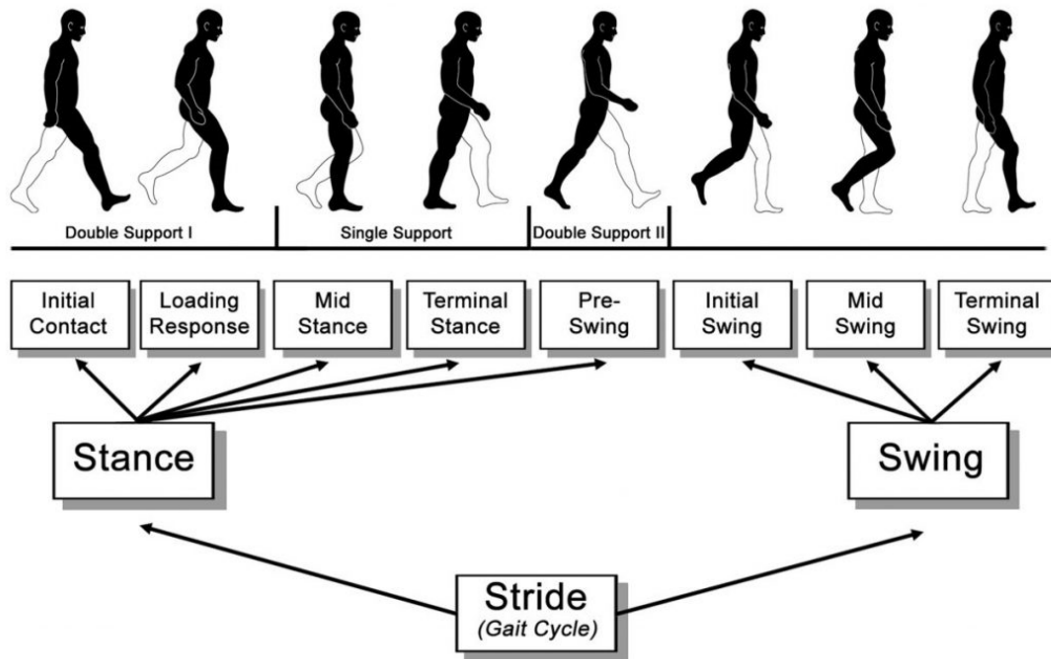


Fig. 2.1: Division of the GC.

Each stride involves an ever-changing alignment between the body and the supporting foot during the stance phase and an advancement of the limb in the swing phase. These interactions result in a series of movement patterns performed primarily by the hip, knee, and ankle, satisfying a functional demand (Bonney-Mazure and Armand, 2015). Within the GC, eight subphases are identified, each related to a movement pattern executed to achieve a functional goal. The sequential combination of the gait subphases allows the limb to perform three basic tasks: weight acceptance, single limb support, and limb advancement (Perry, 1992).

Below, a detailed description for each of the subphases that comprise the GC is provided .

- **Initial contact** (0 - 2% GC): This period includes the moment when the foot touches the ground and reaches a proper landing position (see Fig. 2.2 (a)). The initial contact is often called heelstrike, since in healthy individuals there is often a distinct impact between the heel and the ground.
- **Loading response** (2 - 10% GC): This is the initial double-stance period. It begins with initial contact and continues until the other foot is lifted to swing (see Fig. 2.2 (b)). The objective of this period is to transfer body weight to the stance leg and maintain the forward tendency of the body.
- **Mid-stance** (10 - 30% GC): In this period, the GC enters the single-limb support period. The stance leg alone supports the whole body weight, and the contralateral leg is in its swing phase (see Fig. 2.3 (a)).
- **Terminal stance** (30 - 50% GC): This period completes single-limb support. It begins with the heel rise and continues until the other foot hits the ground, which accelerates the stance limb advancement (see Fig. 2.3 (b)). Throughout this phase, body weight moves forward from the forefoot.

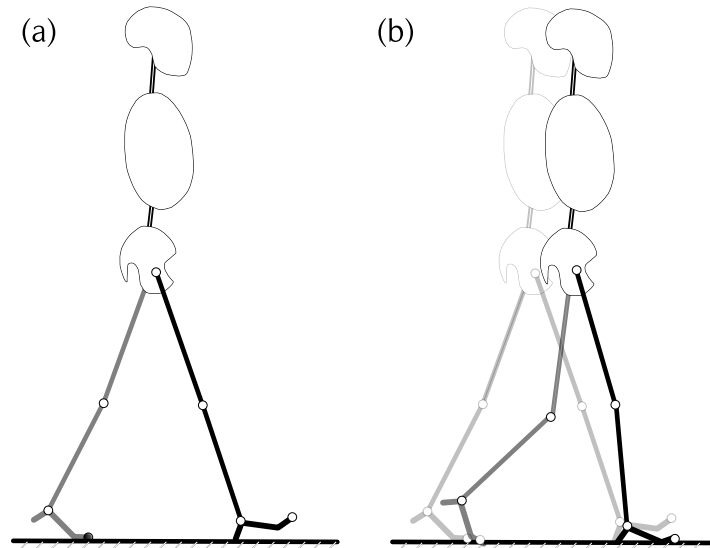


Fig. 2.2: Position of right leg (black) at (a) Initial contact and (b) Loading response.

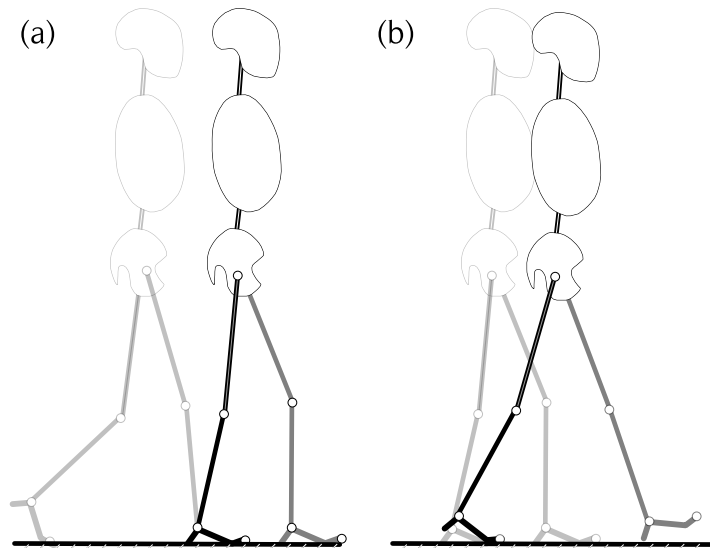


Fig. 2.3: Position of right leg (black) at (a) Mid stance and (b) Terminal stance.

- **Pre-swing** (50 - 60% of the GC): This period is the second double-stance interval. It begins with initial contact with the contralateral leg and ends with the stance leg toe off. Body weight is gradually unloaded from the stance limb to the contralateral side, allowing the toes of the supporting leg to leave the ground (see Fig. 2.4 (a)).
- **Early swing** (60 - 75% of the GC): This period is the beginning of the swing phase. It starts with the toe off and ends when the swinging foot is opposite the stance foot. During this period, the leg moves forward with the ankle dorsiflexed to clear the foot off the ground (see Fig. 2.4 (b)).
- **Mid-swing** (75 - 90% of the GC): This period represents the phase between the moment of maximum knee flexion in the swing and the moment when the tibia reaches a vertical position related to the ground (see Fig. 2.5 (a)). The swing leg continues to move forward as a pendulum due to inertia forces.

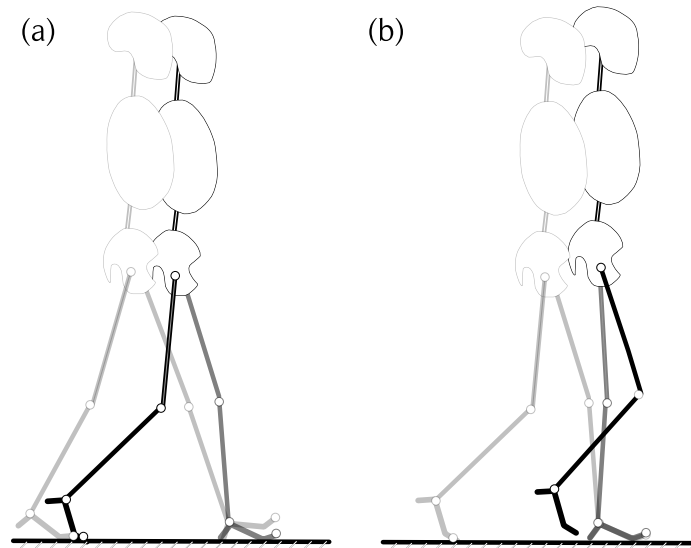


Fig. 2.4: Position of right leg (black) at (a) Pre-swing and (b) Early swing.

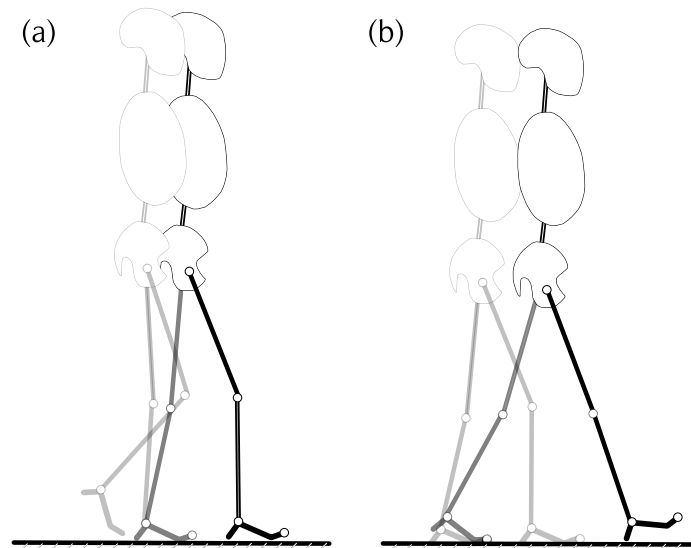


Fig. 2.5: Position of right leg (black) at (a) Mid swing and (b) Terminal swing.

- **Terminal swing** (90 - 100% of the GC): During this period, the tibia continues to advance from the vertical position. The movement of the tibia is decelerated as the knee extends to the next heelstrike (see Fig. 2.5 (b)).

2.2 Spatiotemporal gait parameters

Since the movement patterns performed by humans vary as a function of the gait phase, the quantities related to the movements are instantaneous quantities, i.e., they can be calculated for each infinitesimal instant of time. Similarly, although the normal gait consists of repetitive movements, there are slight variations in the GC even between these repetitions. Therefore, when referring to spatial and temporal quantities that encompass human gait, we refer to average quantities. Observing gait from a spatial and temporal perspective

allows us to measure gait asymmetries related to distance and time between the executed movement patterns (Yusuf et al., 2019). Typical spatiotemporal gait parameters include:

- **Cadence** is the number of steps taken in a given time and is usually reported as steps per min. The mean cadence for children aged 7 years is 143 steps/min, while for adults it is 114 steps/min (Adams and Cerny, 2018). The cadence is calculated using the following expression:

$$C = \frac{Ns}{t} \quad (2.1)$$

where C is the gait cadence Ns is the number of steps in a given time t .

- **Gait velocity**, also called walking speed, refers to how fast the human Center of Mass (CoM) moves forward. The mean speed for children around 7 years is 114 cm/s, while that for adults older than 40 years is 123 cm/s (Adams and Cerny, 2018). The walking speed is calculated using the following expression:

$$GV = \frac{SL \times C}{60} \quad (2.2)$$

where GV is the gait velocity in cm/s, SL is the stride length in cm and C is the gait cadence in step/min.

- **Stride time** is the time measured between two consecutive heel strikes of the same leg (see Fig. 2.6). It is the duration of the GC.
- **Double stance duration** is the period during which the body is supported by two legs. This can also be expressed as a percentage of the GC.
- **Step length** is the length measured parallel to the line of progression of the body from the foot strike of one foot to the foot strike of the contralateral foot (see Fig. 2.6). In normal or nonpathological gait, the lengths of the right and left steps should be symmetric as long as the person walks in a straight line.

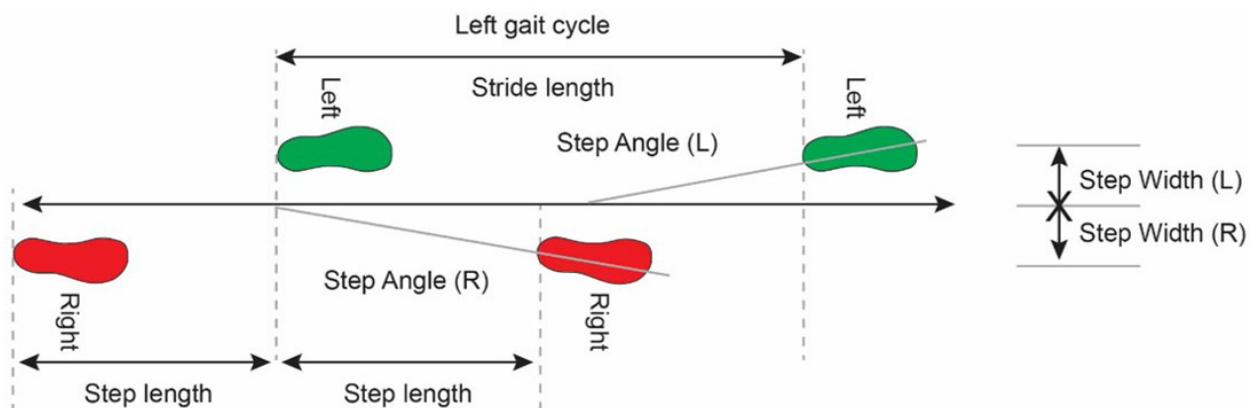


Fig. 2.6: Spatial gait parameters.

- **Step width** is the distance measured between the line of progression of the left foot and the line of progression of the right foot (see Fig. 2.6). This parameter is critical in determining balance problems during gait, as it tends to increase under unstable conditions to increase the base of support.

- **Stride length** is the distance traveled during a complete GC. Each stride length comprises one right step and one left step length (see Fig. 2.6).
- **Foot angle** is the angle between the line of progression and the foot axis (a reference line on the sole of the foot) (see Fig. 2.6). The foot angle is zero when the foot axis is parallel to the line of progression, positive when the foot axis points lateral to the line of progression, and negative when the foot axis points medial to the line of progression. Normal values for this angle are 5° - 10° outward, but toe in can also occur, although less frequently (Adams and Cerny, 2018).

Clinicians initially focus on measuring spatial gait parameters and calculating time-averaged gait parameters as initial steps to assess the 'normality' of a patient's gait, considering their ease of acquisition. However, to achieve a more accurate analysis, additional studies should be carried out.

2.3 Gait kinematics

Gait kinematics is a field of study that delves into the spatial motion described by body parts throughout the GC. It involves the precise measurement and analysis of various parameters, such as joint and segment angles, the spatial position of anatomical landmarks, their velocities, and accelerations. These parameters contribute to a general understanding of the mechanics underlying human locomotion, providing valuable information that extends to clinical rehabilitation, improved sports performance, and ergonomic environmental design.

The subsequent subsection offers a comprehensive description of the angular motion exhibited by the major joints of the lower limb in both the sagittal and frontal planes throughout the GC. The movement patterns presented are based on the book by Simoneau (2010).

2.3.1 Pelvis kinematics

The movement of the pelvis in the sagittal plane is described as short-arc rotations in the anterior and posterior directions about the medial-lateral axis through the hip joints. During gait at normal speed, the amount of anterior and posterior pelvic tilt is small (i.e., a total of approximately 2° to 4°). At initial contact, the pelvis is in a nearly neutral position. From 0% to 10% of the GC, there is a small posterior tilt of the pelvis. The pelvis then begins to tilt forward during the single-limb stance period, reaching a slight anterior tilt of the pelvis just after midstance (30% of the GC). In the second half of the stance phase, the pelvis tilts backward until just after toe off. During the initial and mid-swing (60% to 87% of gait), the pelvis is tilted forward again before beginning to tilt backward in the final swing (see Fig. 2.7 (a)).

In the frontal plane, the pelvis rotates through a total excursion of about 10° to 15° as a result of pelvic-femoral (hip) adduction and abduction in the supporting limb. During weight acceptance in the right lower extremity (i.e., the first 15% to 20% of the GC), the left iliac crest descends slightly below the height of the right iliac crest. From 20% to 60% of the GC, the left iliac crest is lifted by concentric activation of the abductors of the right hip, assisted by a slight displacement of the CoM of the trunk to the right side. Throughout the right swing phase, a similar pattern of controlled initial lowering of the right iliac crest is observed followed by its progressive elevation (see Fig. 2.7 (b)).

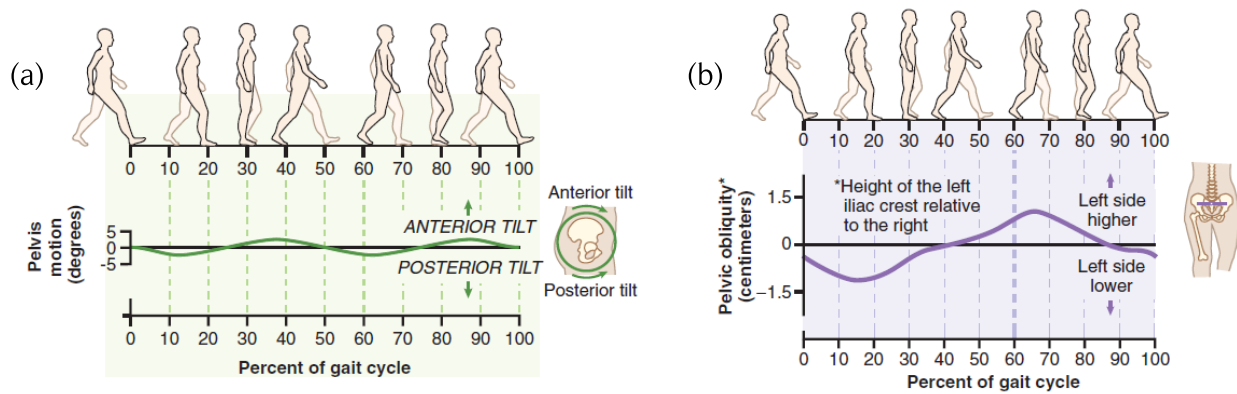


Fig. 2.7: Angular displacement of the pelvis joint in the (a) sagittal, and (b) frontal plane throughout the GC.

2.3.2 Hip joint kinematics

At a typical walking speed, in the sagittal plane, the hip flexes approximately 30° at initial contact. As the body advances over the fixed foot, the hip extends, reaching a maximum extension of approximately 10° before toe off. Hip flexion is initiated during pre-swing, and the hip is at about 0° of flexion at toe off (60% of gait). During the swing phase, the hip flexes further to bring the lower extremity forward for the next foot placement. The maximum flexion (over 30°) is reached just before the initial contact (see Fig. 2.8).

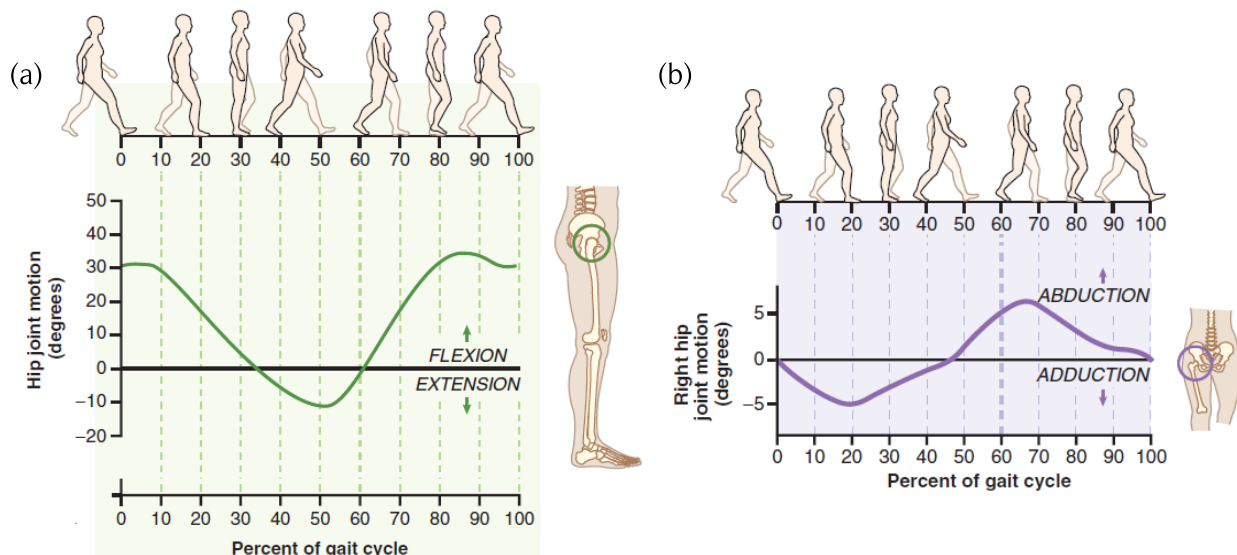


Fig. 2.8: Angular displacement of the hip joint in the (a) sagittal, and (b) frontal plane throughout the GC.

The pattern of elevation and depression of the iliac crests reflects the frontal plane motion of the hips. During the stance phase, this movement in the frontal plane is produced mainly by the movement of the pelvis over the femoral. During the pelvic swing phase, the motion (over the stance limb) combines with the motion of the advancing femur to return the hip to its neutral position in the frontal plane.

2.3.3 Knee joint kinematics

At heel contact, in the sagittal plane, the knee is flexed approximately 5° and continues to flex an additional 10° to 15° during the initial 15% of the GC. This slight knee flexion, controlled by the eccentric action of the quadriceps, serves to absorb shock and accept weight, as body weight is gradually transferred to this lower extremity. After initial flexion, the knee approaches full extension until heel off (between 30% and 40% of the GC). At this point, the knee begins to flex, reaching approximately 35° of flexion at toe off (60% of the GC). The maximum knee flexion of approximately 60° is reached at the start of the mid-swing (73% of the GC). In mid and terminal swings, the knee extends to a near full extension before beginning to flex slightly in preparation for heel contact (see Fig. 2.9 (a)).

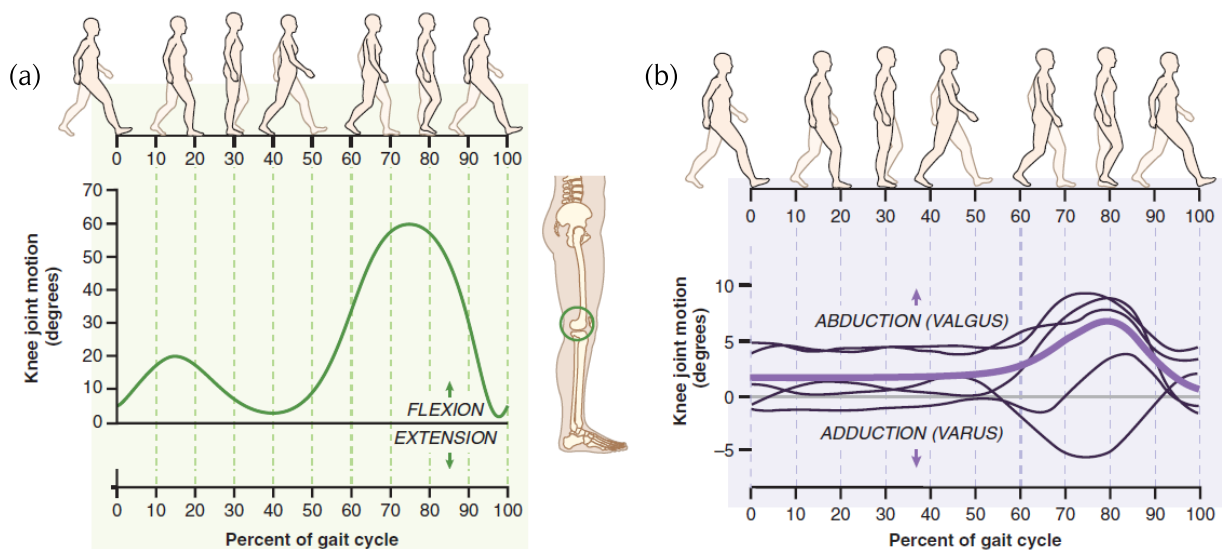


Fig. 2.9: Angular displacement of the knee joint in the (a) sagittal, and (b) frontal plane throughout the GC.

Due to its joint geometry and strong collateral ligaments, the knee is relatively stable in the frontal plane, allowing only a very small amount of angular motion. During the first 80% of the stance phase, the knee exhibits a minimal abduction-adduction motion pattern of less than 3° , while in the last 20% of the stance phase, just before toe off, approximately 5° of adduction occur. The knee normally abducts another 5° during the initial swing phase. Maximum abduction occurred when the knee was close to its maximum angle of flexion. The knee returned to its slightly abducted position before the next heel contact (see Fig. 2.9 (b)).

2.3.4 Ankle joint kinematics

At the ankle, heel contact occurs with the talocrural joint in a slight plantar flexion position (between 0° and 5°). Shortly after heel contact (first 8% of the GC), the foot is placed flat on the ground by an eccentrically controlled plantar flexion movement by the ankle dorsiflexors. Then, during the stance phase, up to 10° of ankle dorsiflexion occurs as the tibia advances over the foot, which is in firm contact with the ground (8% to 45% of the GC). Shortly after heel off (30% to 40% of the GC), the ankle begins to plantar flex, reaching

a maximum of 15° to 20° of plantar flexion just after toe off. During the swing phase, the ankle dorsiflexes back to a neutral position to allow the toes to separate from the ground (see Fig. 2.10).

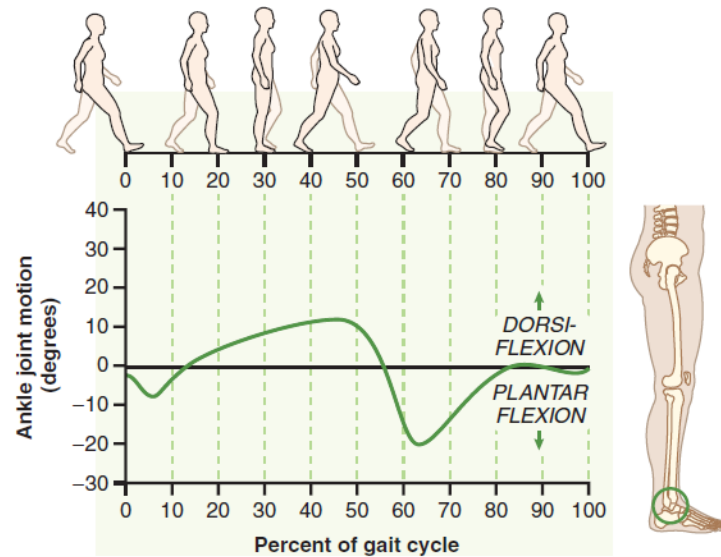


Fig. 2.10: Angular displacement of the ankle joint in the frontal plane throughout the GC.

The primary motion of the talocrural joint is the dorsiflexion-plantar flexion. Although the ankle everts and abducts slightly with dorsiflexion and inverts and adducts slightly with plantar flexion, these secondary movements in the frontal plane are very small and are usually not analyzed.

2.4 Gait kinetics

Understanding the forces involved in gait is crucial in the analysis of normal and pathological movement patterns. Although these forces are not directly observable, they have a direct impact on the observed kinematics. During gait, forces are exerted under the surface of the foot with each step, known as foot forces. Conversely, forces applied to the foot by the ground are called Ground Reaction Forces (GRF) and are equal in magnitude but opposite in direction, according to Newton's Third Law.

2.4.1 Ground Reaction Forces

GRFs are described within a Cartesian coordinate system, with forces expressed along three orthogonal axes: vertical, anterior-posterior, and medial-lateral. The sum of these forces gives a resulting force vector between the foot and the ground (Whittle, 2007).

Vertical GRFs peak twice in each GC, with slightly higher forces at the time of loading response and again at the terminal stance. During mid-stance, the ground reaction forces are slightly less than body weight due to the vertical acceleration of the body's CoM. This fluctuation in force is the result of the need to initially decelerate the downward movement of the body and then accelerate it upward. In terminal stance, the higher ground reaction force reflects the combined push provided by the plantar flexors to reverse the downward movement of the body and prepare for pre-swing.

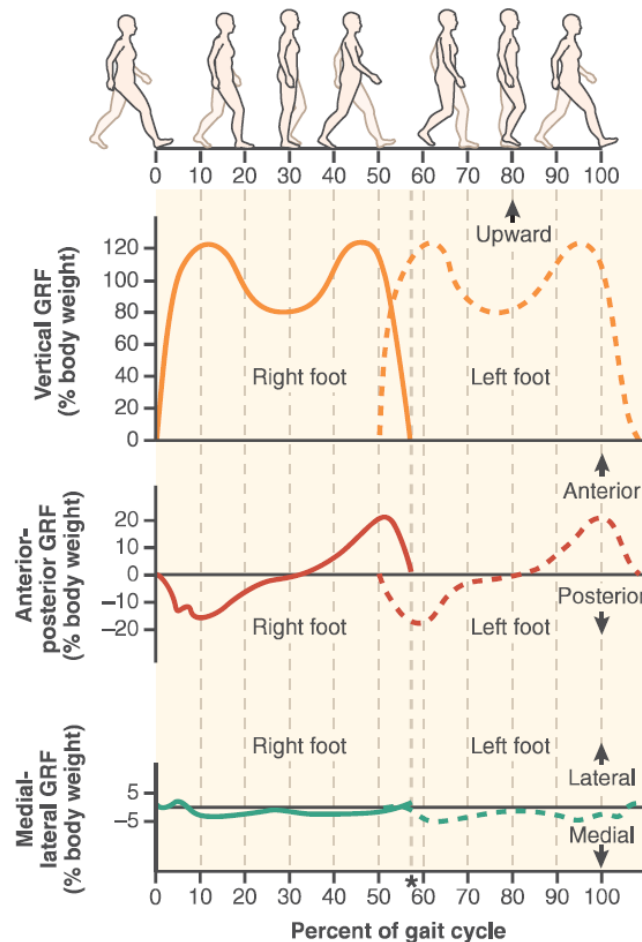


Fig. 2.11: Ground reaction forces (GRF) throughout the GC.

In the anterior-posterior direction, shear forces are applied parallel to the supporting surface. In heel contact, the ground reaction force is directed posteriorly, which requires sufficient friction to prevent the foot from sliding. During terminal stance and pre-swing, the ground reaction force is directed anteriorly, with the foot propelling the body forward by applying a posteriorly directed force to the ground. Friction between the foot and the ground at this stage is crucial for effective propulsion, and inadequate friction can lead to a foot sliding backward without propelling the body forward. The shear forces are influenced by the position of the CoM relative to the foot and are affected by the length of steps, the speed of walking and attempts to accelerate.

Ground reaction forces generate external torques in the joints of the lower extremities during gait. At heel contact, the ground reaction force produces ankle plantar flexion, knee flexion, and hip flexion because the line of action is behind the ankle and knee but anterior to the hip. Internal torques, created by activating specific muscles (such as ankle dorsiflexors, knee extensors, and hip extensors), resist these external torques to prevent collapse of the lower extremity.

Accurate calculations of these forces and torques require the use of an inverse dynamic approach, which involves considering anthropometric characteristics, segment masses, locations of the CoM, mass center inertia matrix, precise body position and motion, and ground reaction forces throughout the GC. Furthermore, the joint power, which reflects the rate of work performed by the muscles and the deformation of the

connective tissues, is derived from the product of the net internal joint torque and the angular velocity of the joint. The positive power of the joint indicates power generation, while the negative power of the joint indicates power absorption, reflecting the activation and stretching of muscles and connective tissues.

2.4.2 Hip joint kinetics

During the initial part of the stance phase, the hip muscles generate a torque in the sagittal plane, resulting in hip extension. This torque serves multiple functions, including accepting the body's weight, controlling the forward momentum of the trunk, and extending the hip joint. As the stance phase progresses, a flexion torque is generated to decelerate the hip extension and then initiate hip flexion prior to toe off. The hip flexion torque is influenced by both passive forces from structures anterior to the hip joint, such as the joint capsule, and the activity of the hip flexor muscles. During the initial swing phase, a small hip flexion torque assists to further flex the hip, corresponding to the concentric activation of the hip flexors. As the swing phase enters its second half, around 80% of the GC, an extensor torque is required to initially slow hip flexion and subsequently initiate hip extension.

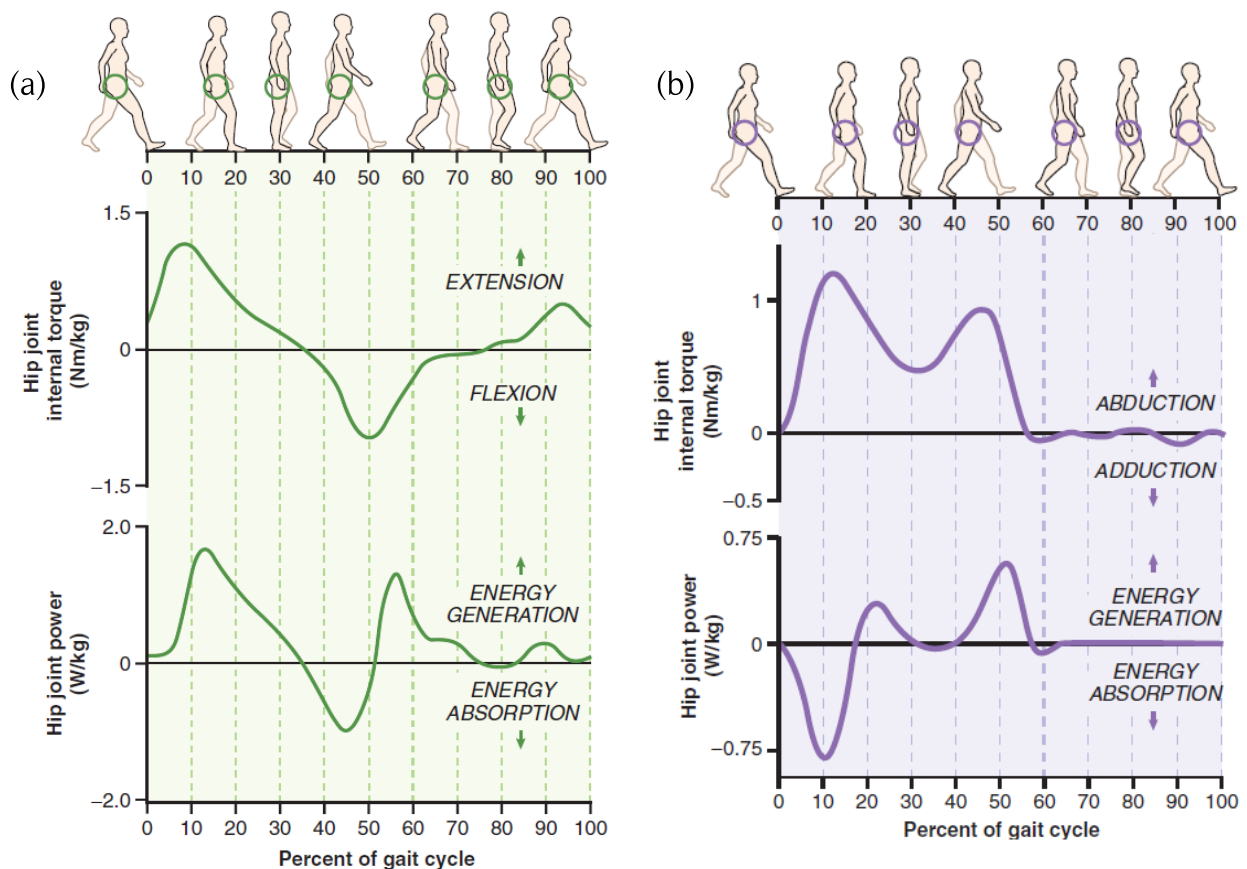


Fig. 2.12: Internal torque and power of the hip joint in the (a) sagittal, and (b) frontal plane throughout the GC.

In the first 35% of the GC, power is generated to support the body, raise the CoM, control trunk movement, and propel the body forward. From there, power absorption occurs until approximately 50% of the GC, reflecting the deceleration of hip extension due to resistance provided by the anterior hip structures

and eccentric activation of the hip flexors. During the pre-swing and initial swing phases, power generation is required to flex the hip. A small amount of energy fluctuation occurs during the latter half of the swing phase, which includes changes in the angular velocity of the hip and the torque necessary to decelerate hip flexion and initiate hip extension.

In the frontal plane, a large abduction torque occurs throughout the stance phase to support the mass of the body that is located medial to the hip joint. Power absorption occurs in the early part of the stance, as the opposite side of the pelvis is initially lowered. This kinematics is controlled by eccentric activation of the hip abductors. Two bursts of power generation are seen in approximately 20% and 50% of the GC, as the contralateral pelvis rises.

2.4.3 Knee joint kinetics

In the sagittal plane, in heel contact, a very brief initial flexion torque (first 4% of the GC) ensures that the knee is flexed to provide adequate knee alignment for shock absorption. A large extension torque needed for the loading response quickly follows this brief flexion torque. This extensor torque continues until 20% of gait has been reached, initially to control knee flexion, then to extend the knee. From 20% to 50% of the GC, there is an internal flexion torque present in the knee despite the knee extending from 20% to 40% of the GC. Knee flexion, in preparation for the swing phase, starts at 40% of the GC, which corresponds to the direction of the internal flexion torque in the knee from 40% to 50% of the GC. However, just before toe off, a small internal extension torque is induced to control knee flexion. In the terminal swing, an internal flexion torque is generated to decelerate the knee extension.

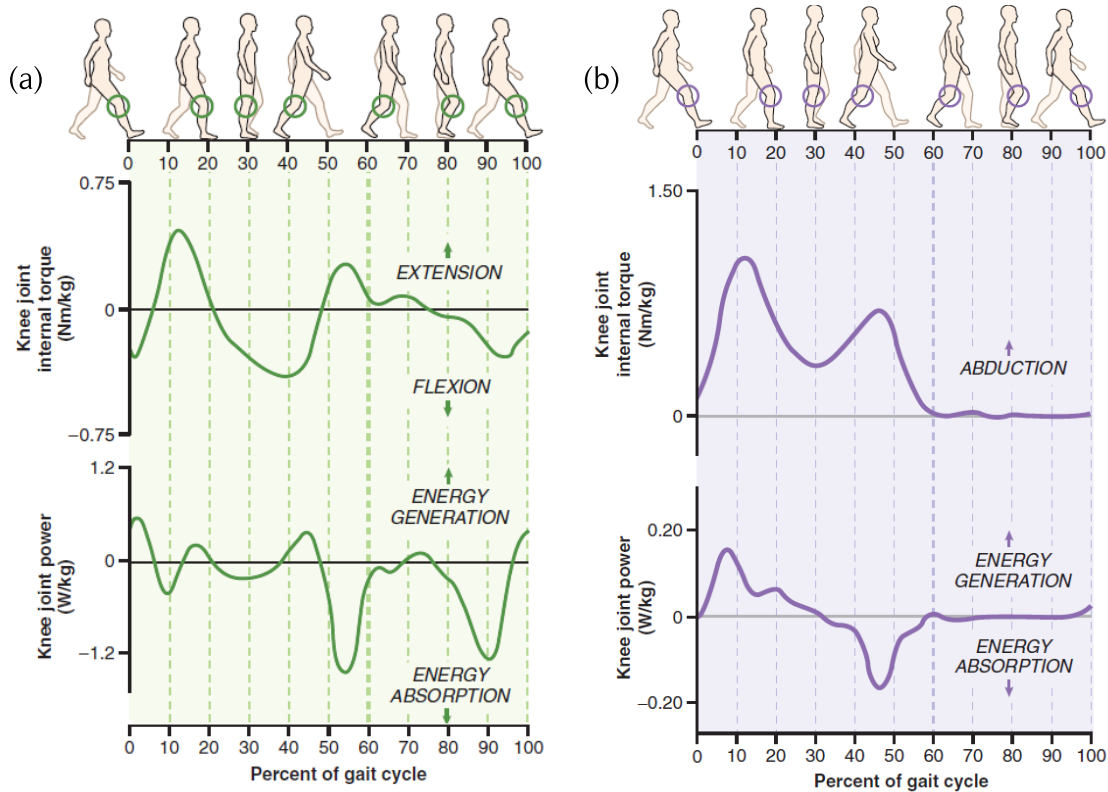


Fig. 2.13: Internal torque and power of the knee joint in the (a) sagittal, and (b) frontal plane throughout the GC.

The power curve for the sagittal plane reflects the action of the musculature surrounding the knee. Power absorption occurs momentarily, reflecting the eccentric action of the quadriceps (from 5% to 15% of the GC). This is followed by another brief moment of power generation, indicating the start of knee extension produced by the continued knee extension torque. Just before toe off, from 50% to 60% of the GC, knee extensors absorb power to control knee flexion. In the second half of the swing phase, the hamstrings absorb energy as the swing limb decelerates, until knee flexion begins just before the next heel contact.

In the frontal plane, during stance, an internal abduction torque at the knee counters the external adduction torque (varus) created by the resultant ground reaction force passing medial to the knee. The internal abduction torque is created by a combination of active and passive structures, including the iliotibial band, the tensor fascia lata, and the lateral ligaments of the knee. Despite the large torques, power values in this plane are very low due to the very small amount of knee movement and hence low angular velocity during stance. However, the pattern of initial energy generation followed by energy absorption suggests a small movement of knee abduction initially followed by a small amount of knee adduction.

2.4.4 Ankle joint kinetics

In the sagittal plane, a small dorsiflexion torque is generated at the ankle immediately after heel contact. This torque serves to control the movement of plantar flexion generated by the application of body weight on the calcaneus. A plantar flexion torque prevails throughout the rest of stance, initially to control the tibia that advances over the foot, then to plantar flex the ankle at push off. During the swing, there is a very small dorsiflexion torque to keep the ankle dorsiflexed and to remove the toes from the ground.

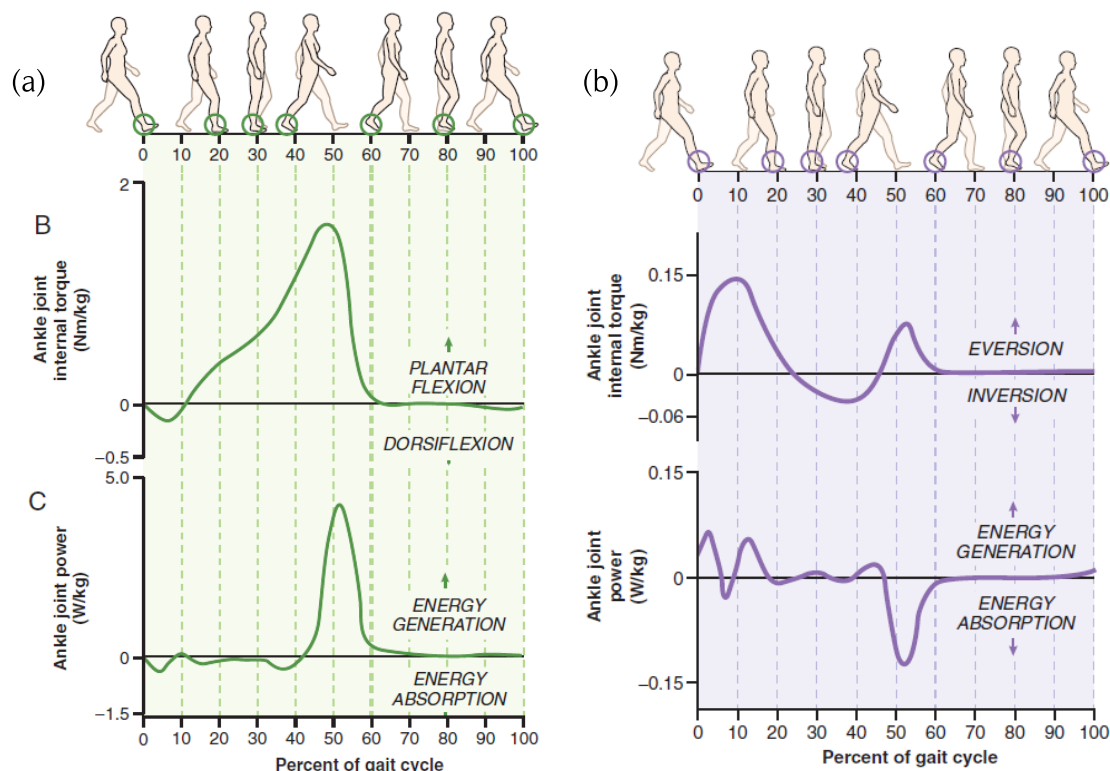


Fig. 2.14: Internal torque and power of the ankle joint in the (a) sagittal, and (b) frontal plane throughout the GC.

In the sagittal plane, power is absorbed just after heel contact as a result of the muscular deceleration of ankle plantar flexion. Then some energy absorption occurs until the push off, reflecting the eccentric activation of the plantar flexors at 10% to 40% of the GC, as the tibia slowly advances over the foot. The relatively slow ankle angular displacement (and assumed velocity) at 10% to 40% of the GC explains the small power values. A large generation of energy occurs at push off (from 40% to 60% of the GC) primarily as a result of a concentric action of the plantar flexors of the ankle, but with some contribution (approximately 10% to 15% of the power burst) from the return of the flexors before heel off. This power generation noted at push-off is considered by many (although not all) to be the primary contributor to the forward propulsion of the body during normal gait, with gravity and the hip extensors also likely playing significant roles.

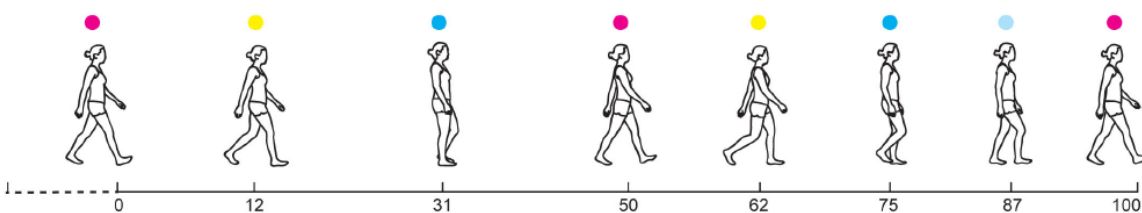
The ankle joint exhibits small torques and significant variation in power values, particularly in the frontal and horizontal planes. In the frontal plane, during the stance phase of gait, a small initial eversion torque is observed from 0% to 20% of the GC. This is followed by an inversion torque that occurs between 20% and 45% of the GC. Lastly, just before toe off, there is a smaller eversion torque. In the horizontal plane, there is the presence of an external rotation torque throughout the stance phase.

Chapter 3

Overview on kinematic gait analysis

3.1 Introduction

Human locomotion assessment is performed using gait analysis studies. In these studies, individuals are subjected to tests in which medical personnel collect information about their body movements while walking. In a medical context, gait analysis helps clinicians and researchers identify structural and functional limitations (diagnosis), determine the severity or nature of a disease or injury (assessment), or predict the outcomes of intervention (Baker, 2006).



	Stance				Swing	
Accomplish	Weight Acceptance		Single Limb Support		Swing Limb Advancement	
Phases	IC	LR	MSr	TSr	PSw / ISw	MSw / TSw
Ankle	☐ Fore Ft.	☐ Foot Slap	☐ Early Heel Off	☐ No Heel Off	☐ Toe Drag	
	☐ Flat Ft.		☐ Inadequate DF		☐ Inadequate DF	
	☐ Abb HC		☐ Excess DF		☐ Contralateral Vault (PF)	
	☐ Inad. DF	☐ Inad. PF				
Calc	☐ Excess Iv.	☐ Excess Iv.	☐ Excess Inversion		☐ Excess Inversion	
	☐ Excess Ev.	☐ Excess Ev.	☐ Excess Eversion			
Toes			☐ Excess IP Flexion (Clawed)			
			☐ Inadequate MTP X			

Fig. 3.1: JAKC OGA form: ankle, foot, and toe deviations. Gait analysis form developed by Pathokinesiology Laboratory and Department of Physical Therapy (2001).

The simplest approach to assess human locomotion is by using Observational Gait Analysis (OGA). In OGA, the physician observes the patient's gait to identify specific deviations from a normal pattern, which

may be related to functional deficits that the patient may be experiencing (Rathinam et al., 2014). Data acquisition techniques used in OGA consist of visual inspection of the patient, through the eyes of the clinician or through the analysis of photos and videos (Adams and Cerny, 2018). The gait characteristics evaluated in the OGA include primarily walking speed, cadence, stride length, as well as temporal characteristics such as the duration of the gait phases and the angular displacement of the lower limb segments. A functional score is generated from the patient's recorded data and compared to previously established performance levels or normative data, which characterize the full spectrum of a specific population (same age group, same gender, etc.). Deviations in recorded data from reference values are usually documented using gait analysis forms (see Fig. 3.1). These forms are made up of check-off options to identify both major and minor deviations (Harris and Wertsch, 1994). They also include the measurement of walking speed, stride length, and cadence with normative gender and age comparisons.

Once gait deviations are documented, a problem-solving approach is initiated to determine the probable causes of each deviation and to differentiate the patient's deficits. The clinician postulates hypotheses about the probable causes of deviations and performs additional examinations based on these hypotheses. In the next step, the results are evaluated with the input of the patient and his family to identify deficits, limitations in activity, and restrictions in participation. Finally, together with the patient and his family, the clinician establishes realistic treatment goals and intervention plans (Hulleck et al., 2022). The main advantage of using OGA is the fact that there is no requirement for specialized equipment or a specific location. However, as a primarily qualitative method, it is relatively subjective in nature, suggesting poor validity, reliability, sensitivity, and specificity (Toro et al., 2003). In addition, its efficacy is largely dependent on the experience of the physician responsible for observing the patient's movement.

Gait analysis plays a crucial role in the assessment of individual conditions, particularly in the elderly population. It involves evaluating various gait characteristics and their connections to gait deficiencies and neurological diseases. Poor balance, slower pace, and reliance on both feet for support are among the initial symptoms observed in certain neurological disorders (Stolze et al., 2002). For example, individuals with multiple sclerosis exhibit altered gait patterns characterized by shorter steps, reduced walking speed, and increased cadence. Osteoporosis, a systemic disease characterized by reduced bone mass, is closely related to gait and balance impairments (Abbass and Abdulrahman, 2014). Fragile bones and an increased risk of fractures are common in people with osteoporosis. Furthermore, given that falls are a primary risk factor for the elderly, assessing gait quality becomes crucial for early diagnosis and intervention (Howcroft et al., 2017). Gait analysis extends beyond the assessment of the elderly population, as gait disorders affect a considerable portion of the global population. Neurodegenerative diseases such as multiple sclerosis, amyotrophic lateral sclerosis, and Parkinson's disease pose significant challenges and require precise gait analysis for diagnosis and treatment (Cicirelli et al., 2022). Furthermore, gait abnormalities manifest in various other conditions, including myelopathies, spinal amyotrophy, cerebellar ataxia, brain tumors, cranioencephalic trauma, neuromuscular diseases, cerebrovascular pathologies, certain types of dementia, heart disease, and the natural aging process (Muro-de-la Herran et al., 2014).

To perform a comprehensive gait analysis, quantitative methods and specialized tools are used. Instrumented gait analysis, conducted in gait laboratories, involves objective measurement of gait using kine-

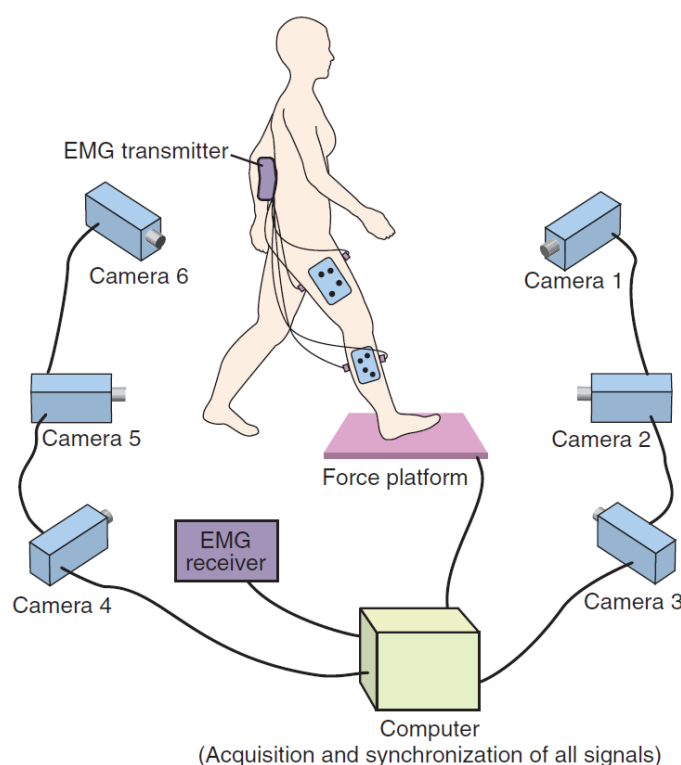


Fig. 3.2: Instrumentation used for gait-related data acquisition in a typical gait laboratory.

matic, kinetic, and EMG signal analysis (see Fig. 3.2). In kinematic analysis, measurements such as displacements, velocities, and accelerations of selected body elements, linear and angular, are collected by tracking body marks and their pattern is analyzed. Kinetic analysis aims to calculate the forces responsible for body segment movements, utilizing a Newtonian or Lagrangian approach. Finally, analysis of Electromyography (EMG) signals provides information on muscle activity during different phases of gait.

Gait analysis, especially when performed with instrumented methods, has proven to be a valuable tool in various medical and healthcare applications. It helps identify and monitor gait abnormalities (Mehrizi et al., 2019), evaluate the effectiveness of orthotic or prosthetic interventions (Healy et al., 2018), assess the risk of falls (Chen et al., 2016), and determine appropriate rehabilitation treatments. Clinicians and researchers rely on instrumented gait analysis to assess gait in different populations, from children with cerebral palsy to individuals with spinal cord injuries, stroke rehabilitation patients, and those with neurodegenerative and motor disorders such as Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis, and Huntington's disease (Cicirelli et al., 2022). Furthermore, the integration of machine learning techniques into gait analysis has shown promising results. Gait pattern classification algorithms automatically identify and categorize different gait patterns, providing a more precise understanding of individuals' conditions. Cluster analysis techniques help identify subgroups of patients with similar gait abnormalities, facilitating disease progression monitoring and personalized treatment strategies. The use of machine learning methods in gait analysis improves diagnostic capabilities, allows for personalized interventions, and allows for monitoring of treatment outcomes.

In general, the combination of quantitative methods, instrumented gait analysis, and machine learning techniques offers valuable insights into gait analysis, leading to more accurate assessments, classification of

gait patterns, and identification of subgroups with similar gait abnormalities. These advances contribute to a better understanding of neurological diseases and optimization of treatment approaches for people with gait impairments. This section provides an overview of the methods and instruments used to collect kinematic data in gait analysis. In addition, it discusses relevant literature studies that focus on utilizing gait analysis to assess individuals' conditions.

3.2 Methods for kinematic data acquisition

3.2.1 Vision-based methods

With the advancement in computer vision and image processing, vision-based methods have become powerful tools for gait analysis, offering objective and quantitative assessments of human walking patterns. These methods use cameras and sophisticated algorithms to capture, analyze, and provide valuable insight into an individual's movement patterns, balance, and potential abnormalities. Vision-based methods for gait analysis can be categorized into two primary approaches: model-free and model-based.

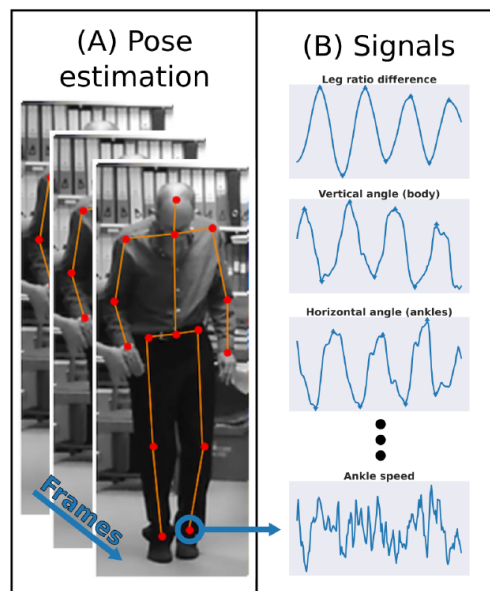


Fig. 3.3: Use of vision-based methods in gait analysis. (A) Key body points tracked in each frame. (B) Extracted kinematic signals (Rupprechter et al., 2021).

In model-free methods, vision-based systems analyze gait patterns without relying on a predefined model of human movement. Relevant parameters, such as step length, cadence, joint angles, and temporal characteristics, are extracted from video footage of individuals walking. By applying image processing techniques such as silhouette extraction, motion tracking, and pattern recognition, model-free approaches can provide detailed information about an individual's gait characteristics (Nithyakani and Ferni Ukrit, 2020). This approach is particularly useful in scenarios where accurate models of human movement are challenging to develop or unnecessary for analysis. On the other hand, model-based approaches in gait analysis rely on predefined models or templates of human movement. These models can be derived from prior knowledge or constructed from a database of gait samples. The vision-based system compares the acquired video data

with the model to estimate the gait parameters. Model-based approaches are often more precise and accurate in capturing detailed movement information, such as joint angles and segmental motion. However, they require careful model calibration and may be sensitive to variations in walking conditions and subject-specific characteristics.

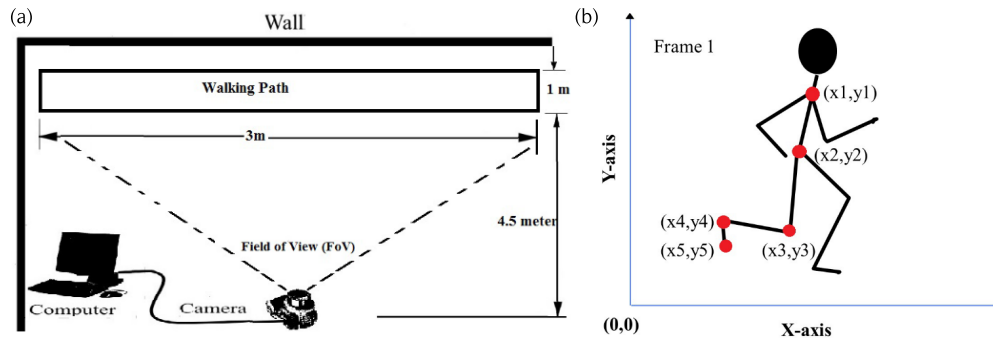


Fig. 3.4: Use of optical-based methods in gait analysis. (a) Laboratory setup for the 2D optical gait parameter analysis system. (b) Marker locations at anatomical points in the subject body (Prakash et al., 2015).

One popular method within vision-based gait analysis is optical-based gait analysis. This approach uses multiple cameras to capture synchronized video footage of an individual walking from different viewpoints. By triangulating the position of body landmarks, such as joints or markers, in captured images, optical gait analysis can reconstruct the three-dimensional trajectory of the body segments during walking. This allows for a precise estimation of joint angles, segmental motion, and other kinematic parameters. Optical gait analysis systems can employ various camera configurations, including monocular cameras, stereo cameras, and RGB-D cameras. Monocular cameras capture color video footage and are commonly used for gait analysis in clinical and research settings due to their affordability and simplicity (Nithyakani and Ferni Ukrit, 2020). Stereo cameras, with two synchronized cameras, enable depth perception and can provide more accurate 3D reconstructions of gait. RGB-D cameras, such as Microsoft Kinect or Intel RealSense, combine color imaging with depth sensing, allowing detailed analysis of gait in three dimensions (Khokhlova et al., 2019).

Marker-based vision systems utilize cameras and markers to aid in human detection, commonly used in controlled environments like research labs. On the other hand, markerless vision-based systems rely solely on cameras to capture video of human gait, eliminating the need for markers, allowing for noninvasive and more naturalistic gait assessments (Verlekar et al., 2018). The image processing techniques are then applied to extract essential parameters for gait analysis.

The benefits of vision-based methods for gait analysis are numerous. They offer objective and quantitative assessments, allowing for more accurate monitoring of changes in gait over time. These methods facilitate large-scale data collection and analysis, allowing researchers to study gait patterns in different populations and conditions. Moreover, vision-based gait analysis can be performed in real-world settings, providing a more ecologically valid evaluation of an individual's walking ability.

3.2.2 Wearable devices

Wearable technology is any kind of electronic device designed to be worn on the body of the user. Among these technologies, wearable sensors are widely used for healthcare monitoring due to their portability, comfort, and relatively low cost. Advancements in micromachining, wireless communication, and network technology have resulted in the development of increasingly miniaturized wearable technology that is capable of being easily integrated into small cases, offering convenience to users without compromising performance. In contrast to motion trackers typically used in laboratories for kinematic data acquisition, wearable devices are well suited for telemetric gait analysis. Accelerometers, gyroscopes and magnetometers, either used individually or combined as a Inertial Measurement Unit (IMU), have become a popular alternative to expensive and rigid laboratory methods for quantifying gait patterns. However, the effectiveness of this approach can be hindered by the incorrect placement of sensors on the limbs and limitations on natural body movements if the sensor size is not sufficiently small (Phinyomark et al., 2018).

MEMS inertial sensors

Microelectromechanical systems (MEMS) integrate mechanical and electrical components into small-scale structures, typically at the micrometer level. These systems are created through a combination of semiconductor and microfabrication technologies that utilize micromachine processing to consolidate electronics, sensors, and mechanical elements onto a shared silicon substrate. The primary components of any MEMS system consist of mechanical elements, a sensing mechanism, and an Application-Specific Integrated Circuit (ASIC) or a microcontroller.

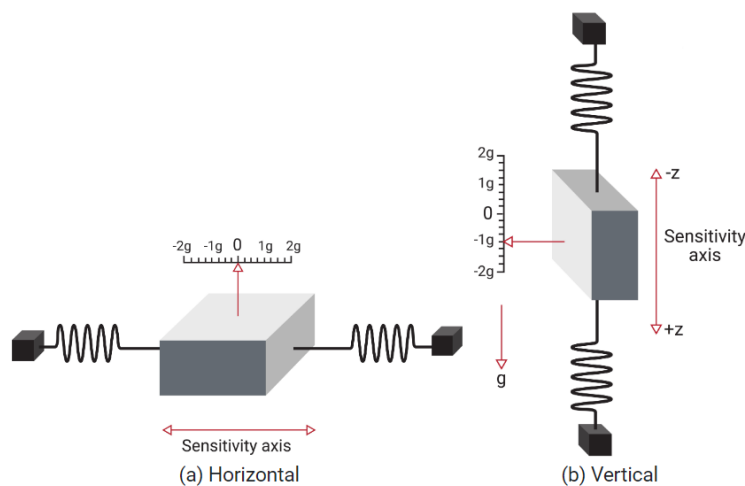


Fig. 3.5: Lumped model of an accelerometer.

Accelerometers are types of inertial sensors that are used to measure linear acceleration along their sensitive axis. The common operating principle of accelerometers is based on a MEMS, which comprises a mass spring system and a position measurement interface circuit. Within the MEMS device, the proof mass is forced to deflect the spring under the action of an inertia force caused by an external acceleration or gravity, following Newton's second law. The position measurement interface unit determines the position of the proof mass, crucial for calculating the inertial force. The method used by the interface unit depends on

the type of accelerometer. Consequently, the acceleration is measured by the displacement changes of the proof mass, and the resulting data are converted into a digital electrical signal through an analog-to-digital converter for further processing.

There are three common types of accelerometer: piezoelectric, piezoresistive, and capacitive. In piezoresistive and piezoelectric accelerometers, the sensing element on which the position measurement interface unit is based is made of piezoelectric or piezoresistive material, respectively. When an inertial force is applied to the device, causing the proof mass to be displaced, the sensing element bends accordingly. This bending generates an output voltage proportional to the proof mass displacement and, consequently, proportional to the applied acceleration. Capacitive accelerometers, on the other hand, enclose the proof mass between two electrodes, and their displacement results in a change in the differential capacitance between the electrodes with respect to the proof mass.

Gyroscopes, on the other hand, are inertial sensors that measure the angular velocity relative to an inertial reference frame. MEMS gyroscopes utilize the measurement of the Coriolis force, an apparent force experienced by objects in motion within a rotating reference frame, whose magnitude is proportional to the angular velocity. To grasp this concept, consider a particle with mass moving on a rotating body relative to an inertial frame. An observer situated in a noninertial frame perceives the particle following a curved trajectory because of the presence of Coriolis and centrifugal forces within that frame. The Coriolis force acts in a direction perpendicular to two quantities: the angular velocity of the rotating frame relative to the inertial frame and the velocity of the particle relative to the rotating frame. Its magnitude is proportional to the speed of the particle within the rotating frame, specifically the component of its velocity that is perpendicular to the axis of rotation.

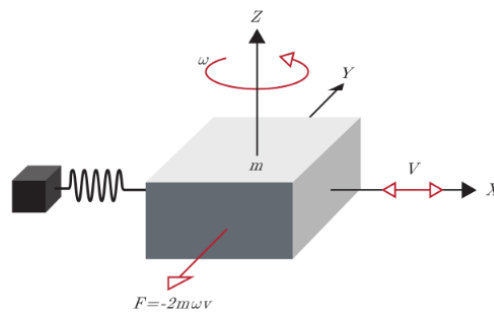


Fig. 3.6: Lumped model of a gyroscope.

MEMS gyroscopes take advantage of the Coriolis effect by utilizing a spring-mass system in which the proof mass resonates with respect to a fixed frame. This fixed frame is rigidly attached to a rotating frame. As a result, the proof mass can only resonate in one direction, causing the Coriolis force to manifest itself in that particular direction, which is perpendicular to the displacement of the proof mass. To measure Coriolis acceleration, the frame containing the spring mass system is attached to an additional frame by springs at 90° relative to the resonating motion. Therefore, the Coriolis acceleration is proportional to the displacement of the proof mass on the horizontal axis.

Finally, a magnetometer is a sensor that is used to measure both the intensity and direction of a magnetic field. MEMS magnetometers operate on the basis of the magnetoresistive effect, whereby, when no

magnetic flux (magnetic field) is applied, current flows directly through a conducting plate. However, when magnetic flux is present, a Lorentz force proportional to the magnetic flux density deflects the current path. This deflection causes the current to travel a greater distance through the plate, resulting in an increase in resistance. In other words, the magnetoresistive effect describes the change in resistivity of a current-carrying ferromagnetic material in response to a magnetic field, with the change in resistance being proportional to the angle of inclination with respect to the direction of the magnetic field. Based on this effect, magnetoresistive sensors can estimate changes in the orientation of a body segment with respect to the magnetic north or the vertical axis.

Flexible goniometers

A flexible goniometer is a type of device used in gait analysis to measure joint angles. It consists of a flexible transducer capable of bending and adapting to the shape of the body segment that is measured. These sensors can be strain gauges, fiber optics, among others.

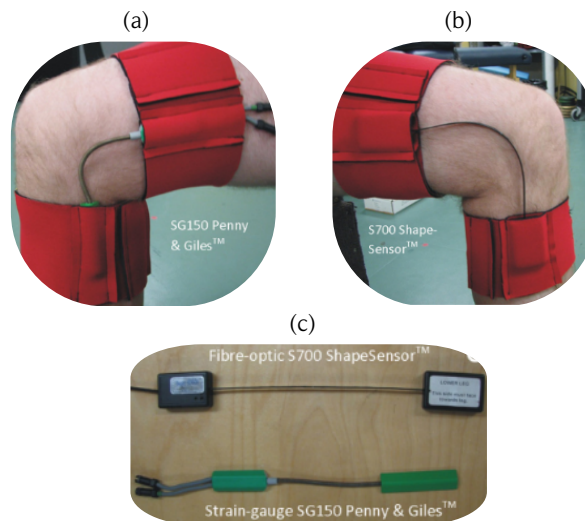


Fig. 3.7: Use of flexible goniometers for measuring knee joint flexion/extension. (a) Strain gauge SG150 Penny & Giles attached to the lower limb. (b) Fiber optic S700 Shape Sensor attached to the lower limb. (c) Image of the goniometers (Mohamed et al., 2012).

There are different types of flexible goniometer available, each with its own principle of operation. Strain gauge-based goniometers utilize the principle of change of electrical resistance in response to strain. They are often placed around a joint or attached to a limb, and as the joint moves, the strain gauge measures changes in resistance, which can be converted into joint angle measurements. Accelerometer-based goniometers use accelerometers to measure the acceleration and orientation of the body segment, and by integrating the acceleration twice, the joint angle can be estimated. Optical fiber goniometers employ the principle of modulating light intensity through bending or stretching of optical fibers, and these changes in intensity are used to determine joint angles.

The flexibility of these goniometers is essential as it allows them to adapt to the contours of the body and be easily attached to different joint locations. This flexibility ensures a comfortable fit for the individual undergoing gait analysis and minimizes interference with natural movement.

3.2.3 Commercial Motion Capture systems

Motion Capture (MoCap) systems have revolutionized the field of gait analysis, providing researchers and clinicians with powerful tools to analyze human motion patterns with unprecedented precision. These systems are typically categorized into two main types: optical and inertial.

Optical MoCap systems utilize high-speed cameras and reflective markers placed on the body to capture and track movement. They are capable of delivering extremely precise measurements of joint angles, segmental motion, and spatiotemporal parameters when performing gait analysis. With the ability to capture motion from multiple camera angles, optical systems offer comprehensive 3D tracking and enable detailed analysis of human motion. In contrast, inertial MoCap systems employ small and lightweight IMUs that contain accelerometers and gyroscopes placed in body segments. These systems excel at portability, ease of use, and real-time motion tracking. IMUs capture the acceleration and angular velocity of body segments, which are then integrated to estimate joint angles and movement trajectories. Inertial systems are less dependent on environmental conditions and can be used both indoors and outdoors. They are particularly suitable for applications that require mobility and freedom of movement, such as sports performance analysis, rehabilitation monitoring, and wearable motion tracking (Muro-de-la Herran et al., 2014).



Fig. 3.8: Setup of a contemporary gait analysis laboratory (BTS Bioengineering, n.d.). (1) Infrared video cameras. (2) Inertial sensor. (3) GRF measurement gateway. (4) Wireless EMG. (5) Workstation. (6) Video recording system. (7) TV screen. (8) Control station.

Both optical and inertial MoCap systems have their strengths and limitations. Optical systems offer superior accuracy and precise measurement capabilities, making them ideal for detailed motion analysis in controlled laboratory settings. However, they require a controlled environment, line-of-sight visibility, and the placement of markers on the body, which can be time-consuming and limit their usability in certain scenarios. On the other hand, inertial systems provide greater flexibility and portability, enabling MoCap in diverse environments and real-world conditions. They are less intrusive and do not require external infrastructure or markers. However, the IMU-based system may have limitations in accurately capturing fine joint angles and may be susceptible to drift over time, which requires regular calibration and integration techniques to improve accuracy.

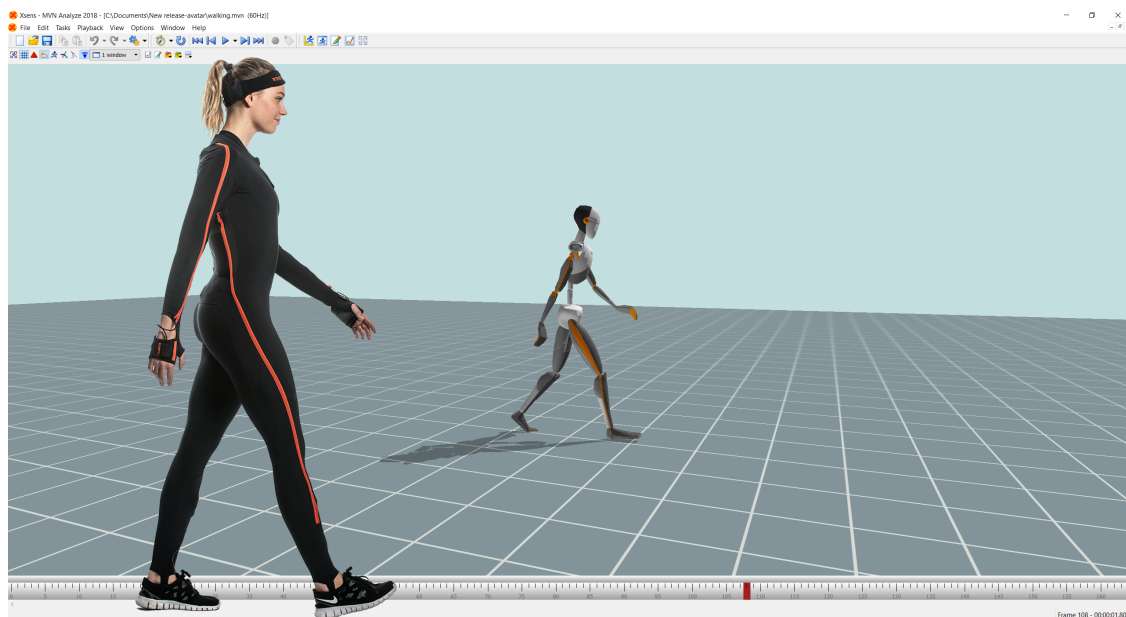


Fig. 3.9: Inertial MoCap suit developed by [Movella Inc.](#) (n.d.).

Commercial MoCaps in gait analysis offer numerous advantages. They provide exceptional precision, allowing for precise measurement of joint angles, segmental motion, and spatio-temporal parameters during gait. This level of detail is instrumental in the diagnosis of gait-related disorders, monitoring progress, and tailoring treatment plans accordingly. Real-time analysis capabilities facilitate immediate feedback and adjustments in clinical and sports performance settings, enhancing the efficacy of therapeutic interventions and enabling more streamlined training and performance optimization. Portability and flexibility are also notable features, as MoCaps can be used in various environments, from clinics to research laboratories, and even outdoors. In addition, user-friendly interfaces and comprehensive software packages simplify data processing, visualization, and interpretation, contributing to a deeper understanding of gait biomechanics.

Several companies develop highly accurate gait analysis systems, including Vicon, Qualisys, and BTS, which are commonly used in laboratory settings. These systems use cameras and other sensors, such as IMUs, force plates, or EMGs, to capture and analyze gait data. They have been used as gold standards in various studies and have contributed significantly to the field of gait analysis. In addition, there are companies that produce smaller-scale products, such as camera sensors or wearable inertial sensor-based systems, catering to specific applications and scenarios. Among the notable examples of commercial MoCap systems, Xsens and OptiTrack have made significant contributions to the field of gait analysis. Xsens (now Movella Inc.) specializes in inertial MoCap technology, combining wireless IMUs with robust biomechanical algorithms. Their systems offer full-body tracking with exceptional accuracy, making them suitable for gait analysis and other biomechanical studies. Xsens stands out for its ease of use, rapid setup, and versatility, making it suitable for gait analysis in various settings, including clinics, sports facilities, and outdoor environments. OptiTrack, on the other hand, is a leading provider of optical MoCap systems. Their camera-based systems offer high-resolution tracking and robust data capture capabilities. OptiTrack systems are widely used in gait analysis, sports science, entertainment, and virtual reality applications.

3.3 Review on kinematic-based gait assessment

Gait analysis based on kinematic data is a powerful diagnostic assessment tool that has seen remarkable advances in recent years, driven by the development of innovative measurement technologies and data mining techniques. Kinematic measures involve analyzing linear and angular displacements, velocities, and accelerations of specific body landmarks. Identification, quantification, and analysis of movement patterns allow the extraction of valuable information for clinical applications, such as functional assessment after treatment or surgery, diagnosis of gait disturbances, and design of gait assistive devices (Molina Arias et al., 2021a) (*author's own research*). This review delves into a selection of notable articles that illustrate the advances made in the field of medical diagnosis based on kinematic gait analysis.

Gait abnormalities encompass complex movement patterns that can involve phenomena such as tremors, partial paralysis, and postural instability. In the field of healthcare, including medical diagnostics, there has been significant interest in methods to classify different types of gait with the aim of recognizing gait anomalies (Kim and Kim, 2021). Gait classification, although it may initially appear to be straightforward, presents a significant challenge due to the substantial variations within gait types, especially among individuals with gait disorders. One of the approaches used for gait classification involves using temporal distances, a method that has long been acknowledged as a prospective diagnostic indicator to distinguish between subjects with neurological impairments (Holden et al., 1986). This approach relies on measuring selected gait features and computing their divergence from reference values obtained from healthy individuals. In cases where the chosen gait features are scalar (such as gait speed, cadence, stride length, etc.), the divergence is essentially calculated by subtracting the values. However, in scenarios that involve multidimensional features, more sophisticated measures of dissimilarity must be taken into account.

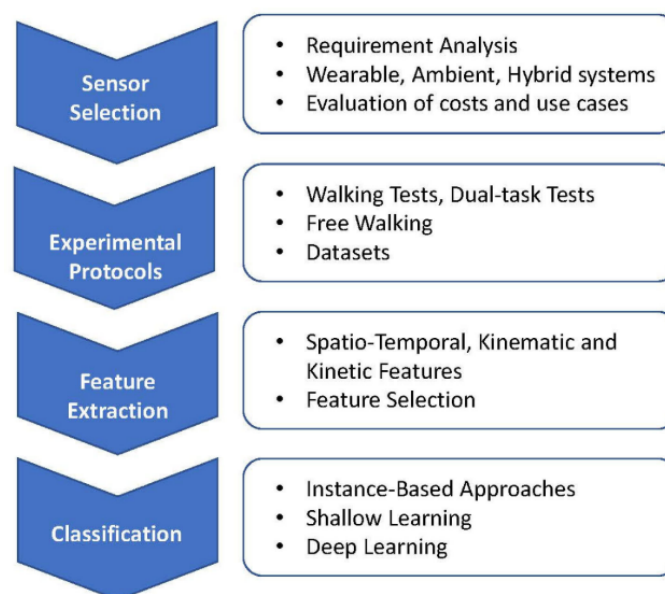


Fig. 3.10: Main stages of instrumented gait analysis in the diagnosis of gait anomalies (Cicirelli et al., 2022).

A notable study conducted by Lee and Shin (2014) illustrates the implementation of similarity measures based on geometric distances within multidimensional gait data. In this investigation, high-dimensional gait

data are collected through experiments using sensors attached to the head, waist, and legs. Then, using distance-based similarity measures, the study unveils intricate patterns and behaviors associated with activities such as walking, standing, and climbing stairs. A parallel study, conducted by [Ramakrishnan et al. \(2018\)](#), introduces a novel gait metric named Comprehensive Gait Asymmetry Metric (CGAM), which employs a modified Mahalanobis distance to evaluate gait parameters extracted from both MoCap and force data during walking. The study effectively highlights the capabilities of CGAM in distinguishing between various gait asymmetries and underscores its potential importance in clinical gait analysis.

The motion patterns of an individual can be visually represented as skeletons by establishing the positions of the measured body landmarks and connecting these points with straight lines ([Wang et al., 2022](#); [Zago et al., 2020](#); [Khokhlova et al., 2018](#)). Analyzing such skeletal structures constitutes a well-established technique for action recognition and holds promise in the diagnosis of gait disorders ([Cavazza et al., 2016](#)). Similarly to the principles of time series analysis, skeletons within a given frame can be compared by applying distance-based techniques. An illustrative example of this methodology can be found in a study by [Khokhlova et al. \(2018\)](#), where a novel method is introduced for the automated detection of abnormal human gait, using an advanced covariance-based approach. This technique involves extracting lower limb flexion angles from skeleton data obtained by Kinect sensors. Subsequently, the covariance matrices are computed on the basis of these angles for each gait sequence. These matrices are then used as features for classification, using both a normal gait-based model and a K-nearest neighbor model. The resulting compact and comprehensive representation of gait provides an effective means of detecting abnormal patterns in gait, all without the need for prior temporal segmentation.

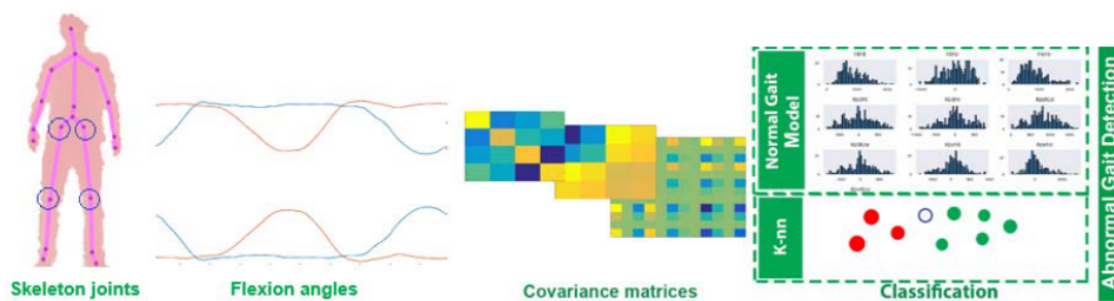


Fig. 3.11: Gait assessment pipeline proposed by [Khokhlova et al. \(2018\)](#).

The use of gait metrics, whether derived from geometric distance calculations or similarity measures, allows for the quantification of the dissimilarity between patterns. This task constitutes the initial phase of a gait anomaly recognition pipeline. After calculating the degree of dissimilarity in relation to a reference pattern representing a healthy state, gait anomaly detection can be performed smoothly by discriminating the most pronounced deviations perceptible to the clinical expert. However, this straightforward approach is less obvious when dealing with considerable volumes of multidimensional data. In such cases, it is more optimal to employ machine learning-based discrimination techniques ([Ferber et al., 2016](#); [Khera and Kumar, 2020](#)).

Extensive research efforts have been ongoing and continue to evolve in the realm of classifying gait patterns through the use of artificial intelligence ([Alfayeed and Saini, 2021](#)). A significant portion of these investigations focus on distinguishing gait patterns specifically related to patients with conditions such as

Parkinson's disease or stroke (Cicarelli et al., 2022; Mannini et al., 2016). Machine learning-based methodologies are used predominantly to diagnose gait disorders, with neural networks (Kidziński et al., 2020; Alharthi et al., 2019), random forests (Bayat et al., 2017), and support vector machines (Khera and Kumar, 2020; Ng et al., 2011) being the techniques used the most. The task of detecting or recognizing a given pathology can be formulated as a binary classification between patients with a health problem and control subjects. However, the possibility of classifying multiple diseases that affect patient gait has also been explored (Wang et al., 2020). In addition, there are alternative approaches that involve modeling an abnormal gait for subsequent analysis using artificial intelligence. An example of this can be observed in the research conducted by Chen et al. (2007), where a combination of hidden Markov models and machine learning algorithms is used to unravel the intricacies of abnormal gait patterns.

Although pattern classification is the predominant approach to the automatic diagnosis of gait disorders, it is important to recognize that this task does not rely exclusively on this method. Clustering is another valid technique, which offers an alternative way to characterize the patterns exhibited by individuals affected by specific gait disorders (Kim et al., 2021). An example of clustering applied to gait discrimination is the study conducted by Chantraine et al. (2022), which significantly illustrates that a dual clustering approach adequately captures the kinematic patterns of the knee in different planes, highlighting the efficacy of this method in gait analysis. Similarly, in the study carried out by Mulroy et al. (2003), patients in the early and late recovery phases after stroke were classified using nonhierarchical cluster analysis, focusing on temporal/spatial and kinematic parameters of walking. Four different clusters of patients were identified at both the initial assessment and the 6-month follow-up, demonstrating the efficacy of the cluster-based approach.

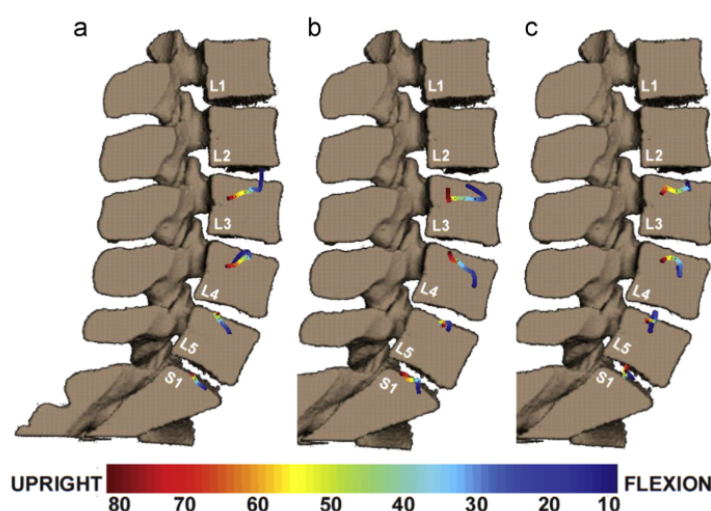


Fig. 3.12: ICR trajectory of intervertebral segments during weight-bearing tasks with (a) 4.54 kg, (b) 9.1 kg, and (c) 13.6 kg. The color map indicates the progression of the extension motion (from flexion to upright) from 10% to 80% of the L2-S1 range of motion. (Aiyangar et al., 2017).

In addition to typical gait kinematic features, parameters derived from acquired kinematic data, such as Instantaneous Center of Rotation (ICR), play an important role in the evaluation of body movements. The ICR provides information on rotational pathways, especially useful in the study of nonperfect revolute

joints, allowing the detection of joint deficiencies. In their study, (Montgomery et al., 1998) proposed the use of the thigh ICR with respect to the shank to measure the actual rotational path of the knee during gait and ultimately detect joint deficiencies and dynamic instability of the joint. Similarly, (Frankel et al., 1971) used the same concept to elucidate the mechanism by which tissue alteration causes mechanical alteration, demonstrating the correlation between pathophysiology and joint disease. Another application of ICR is described by Aiyangar et al. (2017), in which ICR of the lumbar motion segments was calculated during lifting tasks to examine the variations caused by the magnitude of the weight lifted.

Despite interest in gait analysis for the recognition of health problems, the vast majority of studies conducted exclusively on kinematic analysis have focused on analyzing the articular motion of the joints, and to our knowledge, no study has been published that uses ICR as a distinguishing feature of gait. In particular, ICR trajectories within the lower limb segments hold significant promise as a key indicator to assess gait quality and diagnosing gait abnormalities. This concept serves as the focal point of the current thesis, presenting a study with multifaceted implications. From the identification of medical conditions to the development of prostheses or orthoses and expanding to injury prevention and optimization of sports performance, the exploration of ICR trajectories can have wide-ranging implications. In essence, this research contributes to a better understanding of human locomotion, with significant implications for physiological health and rehabilitation.

Chapter 4

Characterization of human gait by means of the ICR trajectories of the lower limb segments in the sagittal plane

4.1 Introduction

As described in Chapter 2, the main elements of the human musculoskeletal system that contribute to generating the characteristic movement pattern observed during gait are the elements that make up the lower extremities. Therefore, the simplest gait models are limited to the study of motion of the lower extremities in the sagittal plane. Despite the intricate nature of the musculoskeletal system, efforts are made to simplify its complexity without overlooking essential elements. In the presented study, the musculoskeletal system is analyzed from a kinematic point of view; this means that the only features considered are movements of anatomical elements of the body, without paying attention to the forces generated by the muscles. The simplifications made are as follows:

- Each lower limb comprises two segments: the thigh and the shank.
- The length of each segment remains constant during its movement (the segments are considered rigid bodies).
- The movements of the segments in each anatomical plane are not correlated with each other; therefore, they can be analyzed independently.
- The location of the CoM of each segment of the lower limb remains fixed with respect to that segment during the entire movement.
- Musculoskeletal joints are considered revolute joints in the analyzed plane.

The movement of lower limb segments in the sagittal plane was analyzed on the basis of the principles of planar kinematics for rigid bodies, following an approach similar to that employed by [Molina and Iwaniec \(2022\)](#) (*author's own research*). To ensure a clear understanding of this analysis, key concepts related to planar kinematics will be introduced, laying the groundwork for the conducted original research.

4.1.1 Principle of planar kinematics

In the plane, a rigid body without imposed constraints has three degrees of freedom corresponding to two translations and one rotation. From their combination, three types of motion can be distinguished and are described below.

A rigid body undergoes **pure translation** if every straight line within the body (line between two particles of the same body) maintains the same direction during motion, and all particles that constitute the body move along parallel trajectories (see Fig. 4.1).

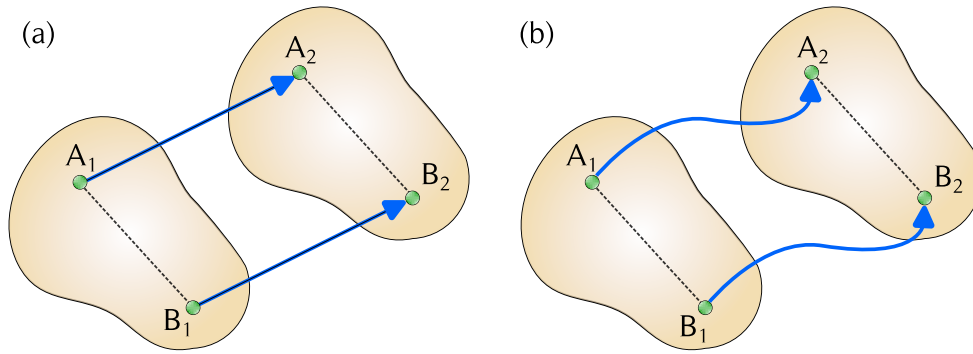


Fig. 4.1: Rigid body undergoing pure translation.

In **pure rotation**, the particles that make up the rigid body move in parallel planes along circles centered on the same fixed point, or, in other words, the body rotates around a fixed point (see Fig. 4.2).

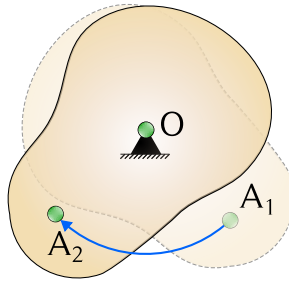


Fig. 4.2: Rigid body undergoing pure rotation around point O .

In the most general case, the body moves with a combination of rotation and translation, known as **general planar motion**. It is important to highlight the fact that the motion of any rigid body undergoing general plane motion can be decomposed into a rotation around a given point and the translation of this point (see Fig. 4.3).

Regardless of the type of planar motion undergone by the body, the position, velocity, and acceleration of the body particles are determined using the same approach, i.e. vectors are used to define the kinematic variables with respect to a chosen frame of reference. If the origin of the reference frame is fixed, it is said to be an inertial reference frame, and the kinematic variables defined with respect to it are absolute.

Consider the rigid body shown in Fig. 4.4, which undergoes a general planar movement. The positions of the points A and B are defined with respect to a reference frame, the origin of which is denoted by the letter O .

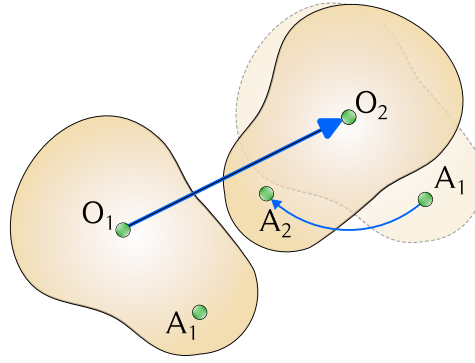


Fig. 4.3: Rigid body undergoing general planar motion consisting of a translation from O_1 to O_2 and a rotation about O .

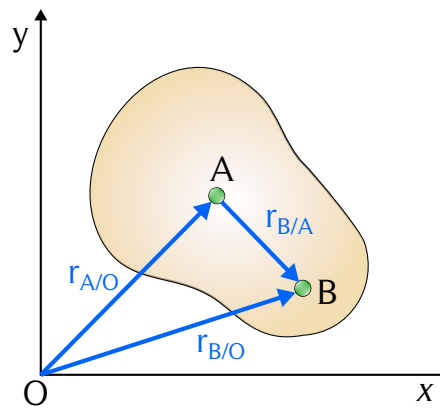


Fig. 4.4: Determination of the position of a particle belonging to a rigid body with respect to an inertial reference system.

The absolute position of B could be obtained directly by measuring the distance from O to B and the corresponding orientation with respect to one of the axes of the reference frame. Another form of determining the absolute position of B is by vectorially summing the relative position of B with respect to A , plus the absolute position of A . In this case, A is usually called the base point and can be any point in the body whose motion is known. Therefore, the absolute position of B is given by the following expression:

$$r_{B/O} = r_{A/O} + r_{B/A} \quad (4.1)$$

where $r_{B/O}$ is the absolute position of B , $r_{A/O}$ is the absolute position of A , and $r_{B/A}$ is the relative position of B with respect to A .

The relative position vector of B with respect to A has a fixed magnitude since it joins two particles of a rigid body (the distance between points belonging to a rigid body is constant), but its direction is subject to change according to the rotational motion of the body. By deriving (4.1) with respect to time, we obtain the expression to compute the absolute velocity of B .

$$v_{B/O} = v_{A/O} + v_{B/A} \quad (4.2)$$

where $v_{B/O}$ is the absolute velocity of B , $v_{A/O}$ is the absolute velocity of A , and $v_{B/A}$ is the relative velocity of B with respect to A .

The term $v_{B/A}$ is equal to the product of the body angular velocity multiplied by the distance between A and B and is perpendicular to the line AB . Therefore, by introducing the angular velocity, (4.2) transforms into the following expression:

$$v_{B/O} = v_{A/O} + \omega \times r_{B/A} \quad (4.3)$$

where ω is the body angular velocity.

To obtain an expression for the absolute acceleration of B , the same approach is followed and the derivation of (4.3) is performed. The resulting expression is presented below:

$$a_{B/O} = a_{A/O} + a_{B/A} \quad (4.4)$$

where $a_{B/O}$ is the absolute acceleration of B , $a_{A/O}$ is the absolute acceleration of A , and $a_{B/A}$ is the relative acceleration of B with respect to A .

The term $a_{B/A}$ consists of two terms that refer to normal and tangential accelerations. Tangential acceleration is the product of body angular acceleration times the distance between A and B and is perpendicular to the line AB , while normal acceleration is the product of body angular velocity squared times the distance between A and B and is parallel to the line AB always pointing towards A . Therefore, introducing the angular acceleration of the body in (4.4), the following expression is obtained:

$$a_{B/O} = a_{A/O} + \omega \times \omega \times r_{B/A} + \alpha \times r_{B/A} \quad (4.5)$$

where α is the body angular acceleration.

At this point, it is important to note that the expressions (4.3) and (4.5) are only valid if the points A and B belong to the same rigid body. If the points are free to reduce or increase the distance between them, then it is necessary to consider the occurrence of additional terms (relative velocity, relative acceleration, and Coriolis acceleration) to describe the effect contributed by a particle moving with respect to a rotating frame.

In pure rotation, it must be satisfied that there is a point belonging to the body with zero velocity around which the body is rotating. Let us consider one more time the movement of the body presented in Fig. 4.3. If the point O is fixed, it means that the rigid body is in pure rotation around the point O . In contrast, if the points O and A have velocities different from 0 it means that the body is undergoing a general planar motion. However, despite undergoing a general plane motion, it is always possible to find another point on the plane that does not necessarily belong to the rigid body around which the body undergoes pure rotation. This point receives the name Instantaneous Center of Rotation (ICR), and knowing its position allows us to calculate the velocity of each point belonging to the rigid body just by knowing the angular velocity of the body and the distance between the given point and the ICR.

For bodies undergoing general planar motion, the position of the ICR varies over time. At any specific time instant, the velocity vectors of the points belonging to the rigid body generate around the ICR a circular field identical to that generated by the pure rotation.

A rigid body that undergoes general planar motion and the position of its ICR at a given moment in time is shown in Fig. 4.5. The position of the ICR is determined by drawing lines perpendicular to the absolute velocity vectors $v_{A/O}$ and $v_{B/O}$, which cross those points, and finding their intersection.

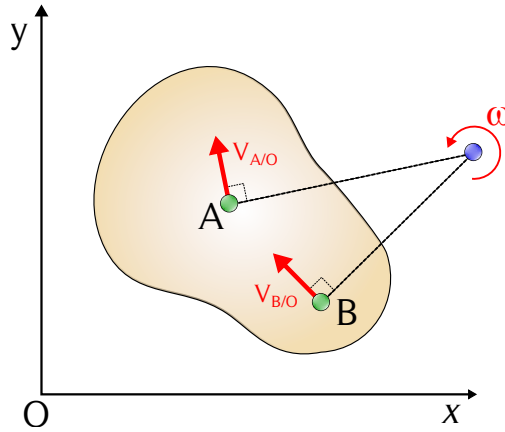


Fig. 4.5: Determination of the ICR position of a rigid body undergoing general planar motion.

The absolute velocity of B could be determined using the absolute velocity of A and the relative position vector between A and B , as described in the previous subsection. Another approach to calculate it is using the position of the ICR. Since the body rotates around its ICR, the absolute velocity is the product of the body's angular velocity and the position vector from the ICR to A . Therefore, the absolute position of B using the concept of ICR is given by the following expression:

$$v_{B/O} = \omega \times r_{B/ICR} \quad (4.6)$$

where $r_{B/ICR}$ is the absolute position of the ICR.

The approach described for calculating the absolute velocity of B consists in transforming the general planar motion of a body into a pure rotational motion. However, the reader should note that, since the resulting motion is always a rotation, it is not possible to obtain an ICR for a body undergoing pure translation. When a body only translates at a given time, its ICR is located theoretically at infinity. This becomes evident since the velocity vectors of all points belonging to the rigid body in question will have the same magnitude and direction, and therefore there is no intersection between the velocity vectors of any pair of points.

To determine the ICR position of a rigid body, two procedures are well known and explained in most books on the kinematics of rigid bodies. The first procedure focuses on determining the absolute positions and velocities of two points that belong to the rigid body at a given instant of time. The second procedure is a graphical method proposed by [Reuleaux \(1876\)](#), which is based on determining the absolute position of two points that belong to the rigid body at two consecutive moments of time. Both procedures are presented in Fig. 4.6 and are described below.

In procedure I, the absolute positions of A and B , as well as the velocity vectors $V_{A/O}$ and $V_{B/O}$ are used for the calculation. If the direction of the two velocity vectors is different, the ICR position is obtained by drawing a perpendicular line to $V_{A/O}$ that passes through A and a perpendicular line to $V_{B/O}$ that passes through B and determining the intersection point. In addition, if the direction of the velocity vectors is the same, the ICR can be found by intersecting the line AB with the line that joins the extremities of the vectors $V_{A/O}$ and $V_{B/O}$. Procedure II uses the position of two different points A and B that belong to the same rigid body in two consecutive instants of time. At time i , we name the two points A_1 and B_1 , while at time $i + 1$ the same points are named A_2 and B_2 . The position of the ICR is determined by joining points A_1 and A_2

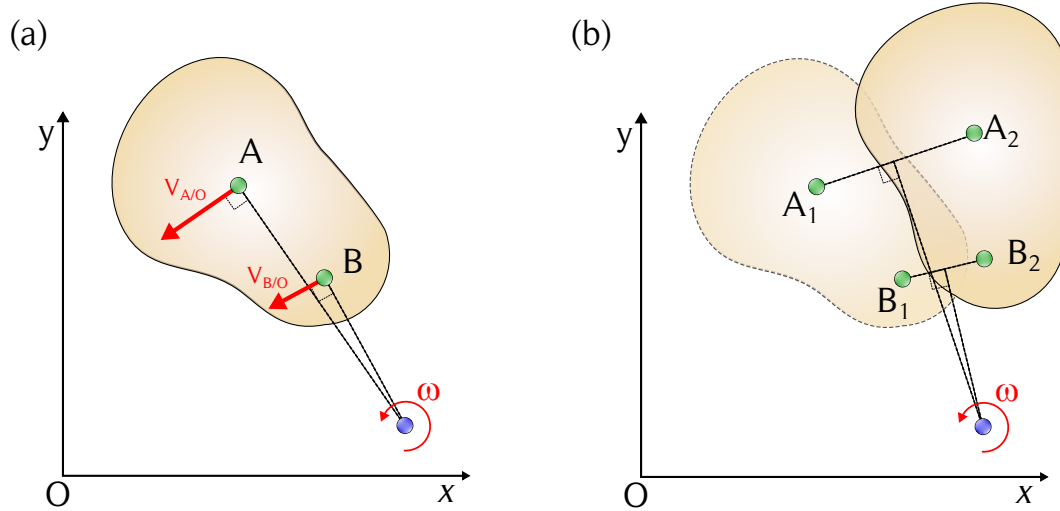


Fig. 4.6: Approaches used for the calculation of the position of the ICR. (a) Technique based on computation of instantaneous velocities. (b) Technique proposed by Reuleaux (1876).

and points B_1 and B_2 using straight lines and finding the intersection of perpendicular lines drawn from the middle points of these two lines.

4.2 Computation of ICR trajectories of the lower limb segments

In each anatomical plane, the segments of the musculoskeletal model undergo general planar motion. As the CoM moves forward with each stride during walking, no body segment has a fixed point of rotation. However, the movement of the body segments can be described as a purely rotational movement around the ICR at any given moment. Determining the ICR position of the body segment over time generates a trajectory that, similar to joint motion, could be a kinematic feature used to extract valuable information about the subject's gait. With this in mind, within the framework of this work we carried out the calculation of the trajectories of the ICR of the human thigh and leg in the sagittal plane and studied its characteristic in the case of subjects without neurological or motor impairments.

The ICR trajectories for the thigh and shank were calculated following the procedure I, i.e., using the absolute position of two points belonging to the analyzed segment and their velocities at a given instant of time. The points considered are the ends of the segments; therefore, to find the position of the ICR of the thigh, it is necessary to acquire the motion of the centers of the hip and knee joints, while to find the position of the ICR of the shank, it is necessary to acquire the motion of the centers of the knee and ankle joints.

Since ICR is the intersection of two lines perpendicular to the velocity vectors of points A and B, the first step that must be taken is to calculate the orientation of lines perpendicular to the velocity vectors, given their components on the x (horizontal) and y (vertical) axes. The orientation of these lines with respect to the horizontal axis is specified by the following expression:

$$\theta_i = \arctan \left(-\frac{V_{xi}}{V_{yi}} \right) \quad (4.7)$$

where θ_i is the orientation of the velocity vector of a given point i , V_{xi} is its horizontal velocity, and V_{yi} is its vertical velocity.

Once the orientation of the two lines perpendicular to the velocity vectors of the points A and B has been calculated, we formulate the equations to find the intersection point between these lines. One of the lines starts from the point A with orientation θ_A and ends at the ICR, while the other line starts from the point B with orientation θ_B and ends at the ICR. From this analysis, we formulate the two equations (on the vertical and horizontal axis) presented below:

$$x_A + l_1 \cos \theta_A = x_B + l_2 \cos \theta_B = ICR_x \quad (4.8)$$

$$y_A + l_1 \sin \theta_A = y_B + l_2 \sin \theta_B = ICR_y \quad (4.9)$$

where X_A and X_B are the positions of points A and B on the horizontal and vertical axes, respectively, l_1 and l_2 are the lengths of the line from A to ICR and the line from B to ICR , respectively, and ICR_x , ICR_y are the positions of ICR on the horizontal and vertical axes, respectively.

Using equations 4.8 and 4.9, the objective is to find the values of l_1 and l_2 , given the positions and velocities of points A and B . Once these values are found, ICR_x and ICR_y are calculated. Given that these equations are nonlinear, the Newton-Raphson numerical method was employed to find the roots of the functions generated when the equations are set to zero.

4.3 Gait-related kinematic data acquisition

4.3.1 Subjects

A total of 15 healthy volunteers without diagnosed gait disorders, with an age range of 28 to 32 years and Body Mass Index (BMI) from 20 to 35 kg/m², were recruited for this study. Among the individuals, eight were men and seven women. Each subject received a full explanation of the purpose and procedure of the study and was subjected to treadmill and overground gait trials. Table 4.1 presents the demographic data collected for each individual.

4.3.2 Instrumentation

Two approaches were used to acquire the kinematic data. The first approach involves the use of a basic but powerful vision-based motion tracking software, which allows one to accurately track the position of markers placed on the main joints of the lower extremities from recorded video. The second approach is the use of a wearable inertial MoCap system, which employs sensor fusion algorithms to provide a comprehensive and highly reliable motion-related data set.

Kinovea[®], which is the vision-based motion tracking software used, is a free 2D motion analysis software that is used primarily to perform biomechanical analysis of athletes. Provides the user with a complete set of tools to track the motion of body landmarks in videos. A notable feature of its operation is its ability to analyze movements without the need to place markers on the subject. However, the developers of the software recommend the use of markers as they greatly facilitate the automatic digitization of body landmarks of interest.

Table 4.1: Demographic information regarding the participants involved in the gait trials.

Subject number	Age	Gender	Height (cm)	Foot length (cm)	Thigh length (cm)	Weight (kg)
Subject 1	28	M	171	26	32	78
Subject 2	27	F	160	25	39	70
Subject 3	30	M	175	29	27	72
Subject 4	29	F	162	24	33	64
Subject 5	26	M	185	32	47	82
Subject 6	31	F	158	26	39	55
Subject 7	31	M	168	28	38	71
Subject 8	31	M	173	25	36	75
Subject 9	30	F	168	28	47	55
Subject 10	30	M	168	28	44	60
Subject 11	27	F	159	28	38	75
Subject 12	29	M	164	28	39	55
Subject 13	31	M	170	28	43	70
Subject 14	29	F	165	26	47	55
Subject 15	30	M	173	28	43	80

In Kinovea, the automatic tracking process consists of calculating the cross-correlation coefficient between the window of interest and the candidate windows in the next frame of the video. When the window of interest is compared to each potential window, several scores are obtained. The highest score represents the best match, indicating that the candidate window closely resembles the original window (window of interest). However, if the score falls below a predefined threshold, it indicates that the target has possibly been missed, and the tracking process may need to be re-calibrated.

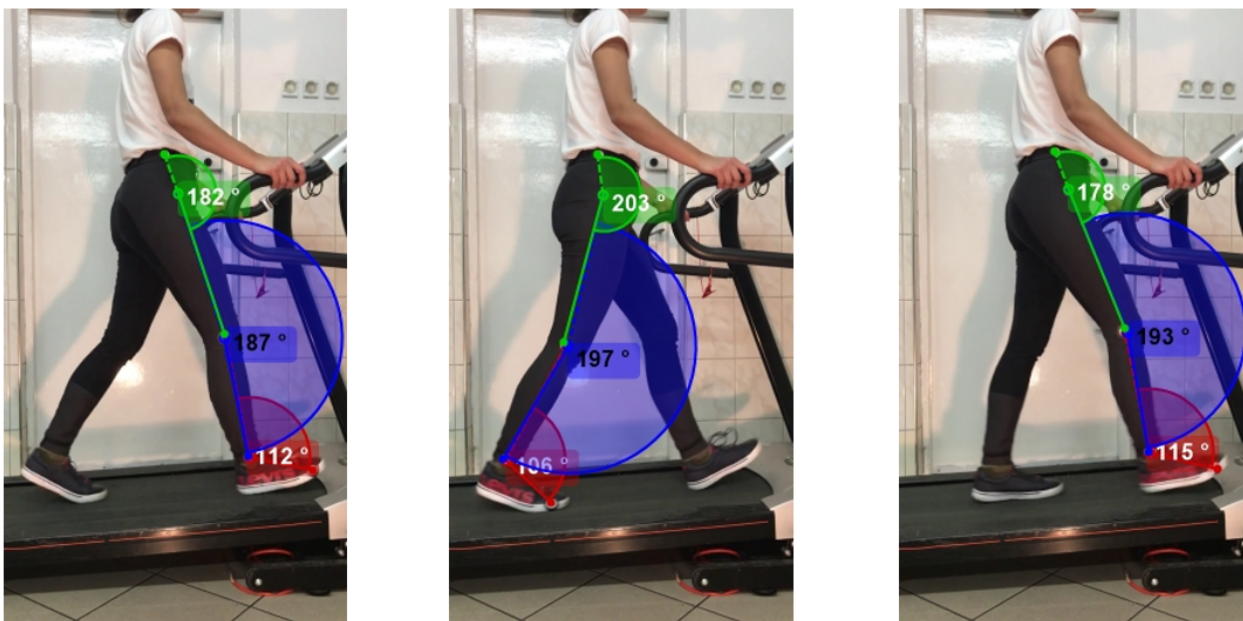


Fig. 4.7: Use of Kinovea for tracking lower limb joint angles.

To record the videos used for the acquisition of kinematic data, GoPro sports cameras are utilized. Specifically, two cameras are used in the process, the GoPro Hero 10 camera and the GoPro Hero 8 Black camera, both configured with identical settings to record the same subject from different planes. This dual-camera setup allows for a three-dimensional view of the subject's movement from one body side, as long as the orientation between the recording planes is not parallel. However, in this research, the idea was to use parallel planes to record the subject's movement in the sagittal plane from the left and right body sides. GoPro cameras offer exceptional video quality, capable of recording at a maximum frame rate of 240 fps. This high frame rate ensures that the finest details are preserved even during fast movements, making the footage ideal for meticulous kinematic analysis. Figs. 4.8 and 4.9 present the technical parameters of the set of cameras.

	Sensor resolution	23 MP
	Video resolution	5.3K (up to 60 fps) 2.7K (up to 240 fps) Full HD 1080p (up to 240 fps) 4K (up to 120 fps)
	Connectivity	WiFi Bluetooth USB
	Image resolution	5568 X 4176
	Sensors	Accelerometer Gyroscope GPS

Fig. 4.8: Technical parameters of sport camera GoPro 10.

	Sensor resolution	12 MP
	Video resolution	4K (up to 60 fps) 2.7K (up to 120 fps) Full HD 1080p (up to 240 fps) QHD 1440p (up to 120 fps)
	Connectivity	WiFi Bluetooth USB
	Image resolution	4240 X 2824
	Sensors	Accelerometer Gyroscope GPS

Fig. 4.9: Technical parameters of sport camera GoPro 8.

The operation of GoPro cameras was controlled remotely through the use of the app Camera tools for GoPro Heroes. This handy app gives users comprehensive control over multiple GoPro cameras with several features that enhance the recording experience. Through the app's interface, the user can easily adjust recording settings, start or stop recording, and even tag significant moments simultaneously on multiple cameras. This feature is essential for conducted research, since synchronization of the recorded videos is required.

The inertial MoCap used for data acquisition was the MVN Awinda system, which is an advanced solution, developed by Xsens, that enables wireless data transmission to a PC or laptop. To capture the movement of the human body, the system uses a setup of 17 motion trackers. These trackers are strategically placed on various parts of the subject's body, including the feet, lower legs, upper legs, pelvis, shoulders, sternum, head, upper arms, forearms, and hands. Each tracker is an independent inertial and magnetic measurement unit composed of a 3D gyroscope, a 3D accelerometer, and a 3D magnetometer. By implementing an advanced signal processing strategy that includes the use of StrapDown Integration (SDI) algorithms, data is sent at a relatively low rate (60 Hz) while preserving sampling accuracy at a much higher rate (e.g., > 1 kHz). Furthermore, wireless data transmission between each tracker and the so-called Awinda station is performed using a patented Xsens protocol. To enhance the user experience, a lycra sweater, customized with zippers on the back and near the sternum, simplifies mounting of the motion trackers at specific locations on the body. The rest of the trackers are attached to the body using body straps. Details about the placement of trackers in the human body are provided in Appendix A.1.

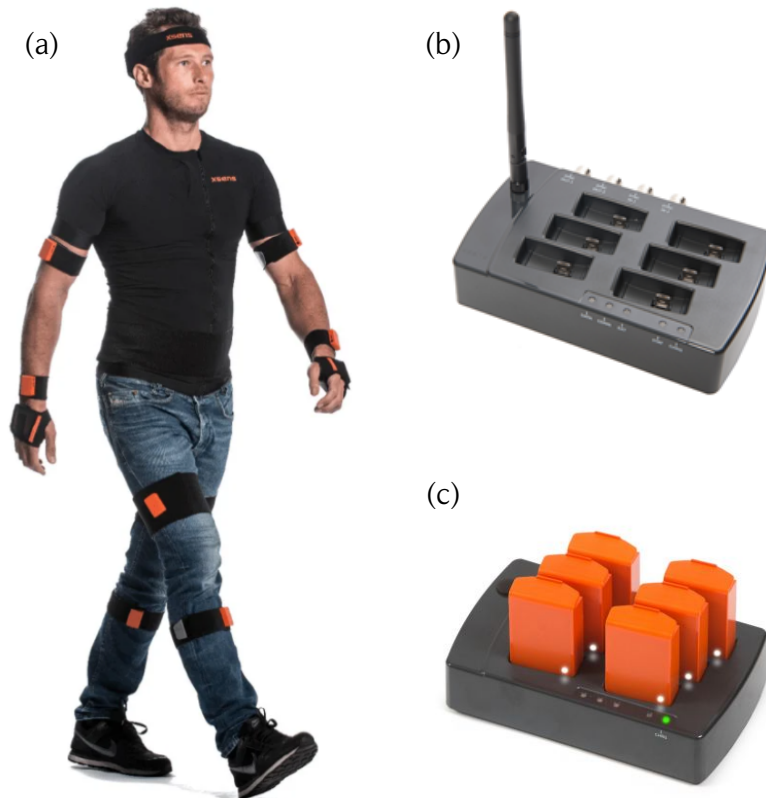


Fig. 4.10: MVN Awinda MoCap system. (a) User wearing the set of motion trackers. (b) Signal receiver. (c) Motion trackers and battery charger.

The core element of the MVN Awinda system lies in its powerful software engine, which plays a crucial role in the motion capture process. This engine combines the raw data from individual motion trackers with an advanced biomechanical model of the human body. The result is a precise determination of the positions and orientations of the body segments. The biomechanical model comprises 23 segments, and their position and orientation is defined with respect to a common, local coordinate frame, providing a consistent and standardized reference for all kinematic quantities.

MVN Awinda system performs an automatic real-time processing frame by frame, progressively advancing over time. In addition, the data are reprocessed in high definition over an extended time interval, which optimizes the estimation of the position and orientation of the body segment and enables highly reliable results.

4.3.3 Experimental protocol

Gait trials were carried out in the Biomedical Signal Processing Laboratory, located in the Faculty of Electrical, Automation, Computer Science, and Biomedical Engineering of the AGH University of Krakow. The gait trials were performed in two scenarios: on a treadmill, where the subjects walked at a controlled speed, and overground, where they walked at their preferred pace.

During gait trials on the treadmill, each subject wore the MVN Awinda system, and GoPro cameras were used to record the subject's movement pattern in the sagittal plane on both the left and right sides. The treadmill was adjusted to maintain a constant speed of 4.2 km/h. The cameras were set to record at 120 fpm and 2.7K video resolution. These cameras were mounted on tripods and aligned so that the lens orientation coincided with the sagittal plane of the recorded subject. The height of the tripod was adjusted to 80 cm and placed at a distance of 150 cm from the subject. Fig. 4.11 shows the treadmill gait trail setup.

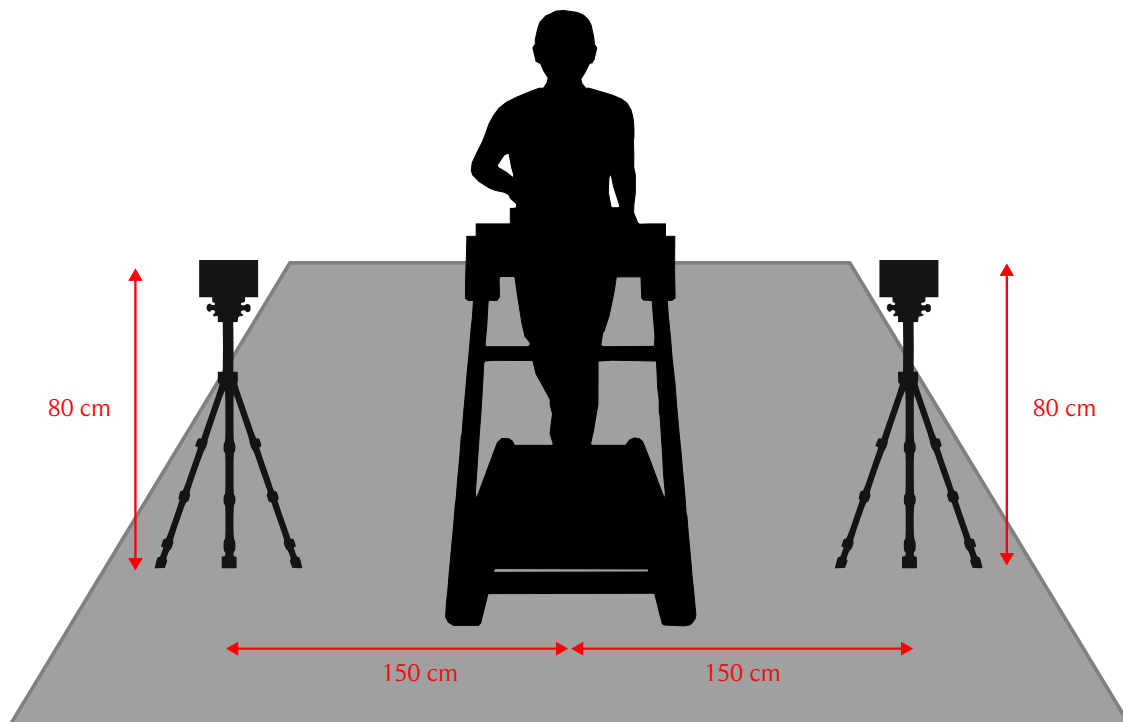


Fig. 4.11: Setup for the acquisition of gait-related kinematic data.

Circular markers printed on standard paper were taped near the hip, knee, and ankle joints so that they could be recorded from the sagittal plane of the subject with the cameras. The purpose of using markers is to create a visually distinctive and easily located landmark in the recordings. Their contrast in relation to the test subject's clothing or skin allows Kinovea software to identify them and automatically track their position in the recorded video. Fig. 4.12 shows the look of the markers used.



Fig. 4.12: Marker used for kinematic gait-related data acquisition using Kinovea.

During the overground gait trials, the subjects wore only the MVN Awinda system. They were asked to walk along a straight 10 m line, while the MVN Awinda system recorded the motion of the trackers. To accurately determine the position and orientation of the individual's body landmarks using the MVN Awinda system, a calibration process needs to be carried out. This process comprises three key steps: Scaling, Segment-by-Segment Calibration, and Axis Definition (Schepers et al., 2018).

Scaling consists of applying a generalized scaling model using a set of input parameters provided by the user. The minimum set of input parameters to be provided by the user is his height and the length of his feet. However, to increase precision, additional measurements can be included, such as arm length, ankle height, hip height, hip width, knee height, shoulder width, and shoulder height. The software engine then utilizes this generalized model to find the best fit based on the provided parameters, resulting in a highly accurate estimate of the person's dimensions and proportions. In this study, only height and foot length were used to scale the model.

Moving on to the segment-to-segment calibration, the alignment of the trackers is measured with respect to the body segments. This measure is obtained by employing a reference pose, in which the segment orientations are assumed to be known at that specific pose. In this stage of the calibration process, the subject is instructed to maintain a still position, adopting an N- or T-pose, and walk a few meters back and forth for a short period of time. Once the segment-to-segment calibration step is completed, the subject is again positioned in an N/T pose with its head facing the forward direction of the measured environment. This setup allows for defining the forward x -direction of the local coordinate system, along with its origin, which is determined at the position of the right heel.

In the case of the vision-based tool, a calibration step is also required. In this case, calibration consists of measuring a distance in the actual setup, as captured by the video, to specify a scale factor for distances within the video. Digital videos consist of frames captured at a fixed frequency, and each frame contains a specific number of pixels determined by the resolution of the camera. Consequently, units of length within the video are represented in pixels, and not in actual physical measurements. Calibration ensures an accurate conversion between pixel measurements and real-world distances, allowing data collected in video recordings to be accurately analyzed and interpreted.

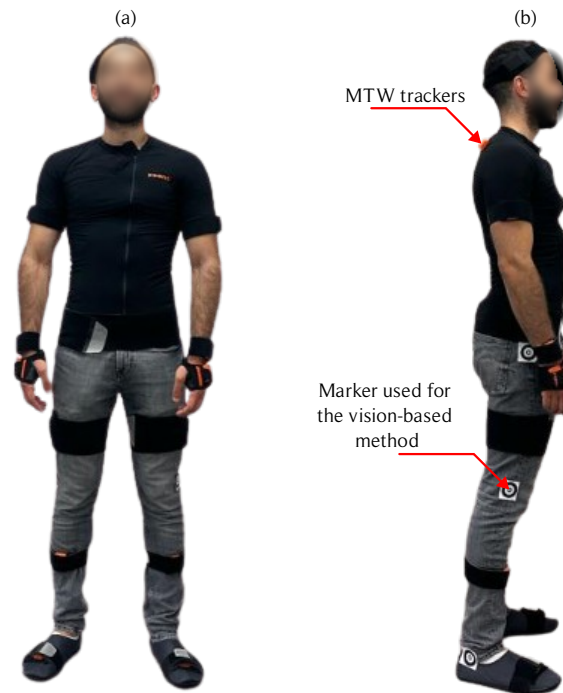


Fig. 4.13: Placement of the trackers and markers. (a) Participant's frontal view, and (b) Participant's lateral view.

The recorded videos of the gait trials of each subject were imported into Kinovea, where markers placed near the centers of the hip, knee, and ankle joints were tracked. At times, due to too fast movements of the subject, the frame resolution was not sufficient for the Kinovea algorithm to automatically recognize a given marker. In these cases, it was necessary to manually identify the marker within the video, which is the main drawback of the method, as it requires a great effort to correctly process the video and obtain the kinematic data. Similarly, the kinematic data acquired through the MVN Awinda system were automatically processed using the MVN analyzer software and exported as .xmls files. The files were imported into Matlab®, where functions for the calculation of the ICR position of the lower extremities segments in the sagittal plane were implemented.

4.4 Results and discussion

From the kinematic data acquired during the gait trials and using the procedure described for the calculation of the ICR, the characterization of the ICR trajectories of the thigh and shank is carried out. In this section, the data acquired for each type of gait trial and their computed ICR trajectories are presented and compared. Because the large volume of data makes it difficult to visualize on the same plot the collected data of different subjects, initially, only the data of an arbitrarily selected subject are presented.

During a gait trial, each subject performs several strides, the number of which depends on several factors, such as the duration of the trial and the speed at which each subject walks. Each stride executed by the subject corresponds to a single GC, and its kinematic characteristics under normal conditions should be common between GCs of the same subject. Due to this fact, it is sufficient to analyze the trajectories acquired during a single GC to draw conclusions about the gait of the subject under study.

The extracted kinematic data are time series, i.e. they are time-related variables. Although the sampling frequency is the same in each conducted gait trial, the number of samples contained in a single GC differs not only between subjects but also between gait cycles of the same subject. Therefore, the segmentation of the gait cycles contained in the analyzed time series must be performed on the basis of the intrinsic characteristics of the time series.

The GC segmentation approach used in this study relies on the identification of the number of peaks within the trajectory of the hip marker on the vertical axis of a single GC. This marker was chosen because its characteristic shape is similar in both; overground, and treadmill gait trials, which ensures consistency in the segmentation process.

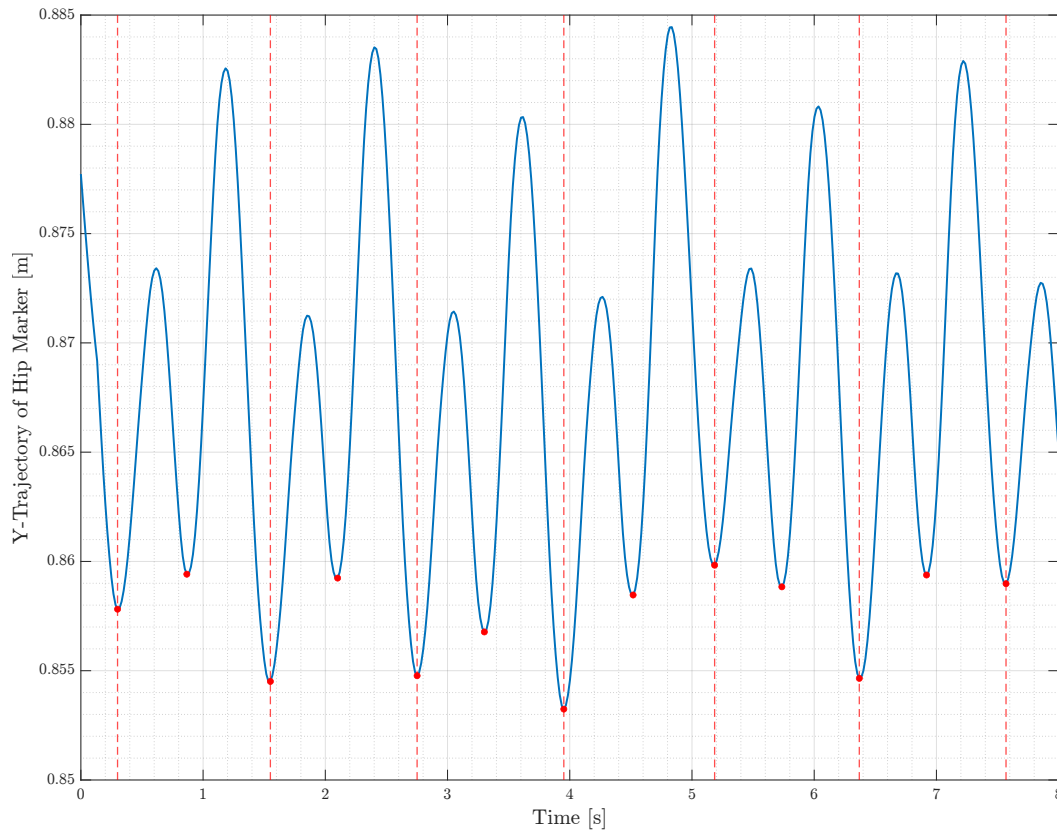


Fig. 4.14: Hip marker trajectory on the vertical axis. The plot shows the principle of the proposed segmentation technique. The red dots are the local minima encountered within the signal, while the red dashed lines indicate the end of one GC and the beginning of the next.

Visual inspection revealed that the trajectory of the hip marker exhibits three distinct lower peaks (local minima) during a single GC. The specific time indices at which these local minima occur were determined by an analysis of the first derivative of the signal. In this analysis, the moments at which the trajectory of the marker changes from a decreasing to an increasing function are identified. These moments correspond to the time of occurrence of the local minima. Once these moments are identified, each acquired time series is segmented into the corresponding GCs by dividing the signal after counting three consecutive local minima on the hip marker trajectory. Fig. 4.14 shows the trajectory of the hip marker along the vertical axis and the resulting signal segments using the explained approach.

To account for the variability of the GC characteristics within a single subject, the set of segmented trajectories is averaged and their standard deviation is calculated. In this way, a representative trajectory of each marker is obtained for each subject. Since each segmented GC has a different number of samples, a linear interpolation is performed so that each of the segments has a number of 200 samples.

Fig. 4.15 shows the representative trajectories of each marker collected in each gait trial for the same subject, and as the reader can notice, the most significant variation between the data acquired in each gait trial is on the horizontal (x) axis. In treadmill trials, the markers exhibit a cyclic back-and-forth motion, whereas in overground trials, the markers advance linearly until the subject comes to a stop. Interestingly, in the case of the data recorded from the treadmill gait trial using the MVN Awinda system, the trajectory of the markers closely resembles that of overground walking. This similarity can be attributed to the principle of operation of the inertial system and the inherent drift error associated with the data integration process.

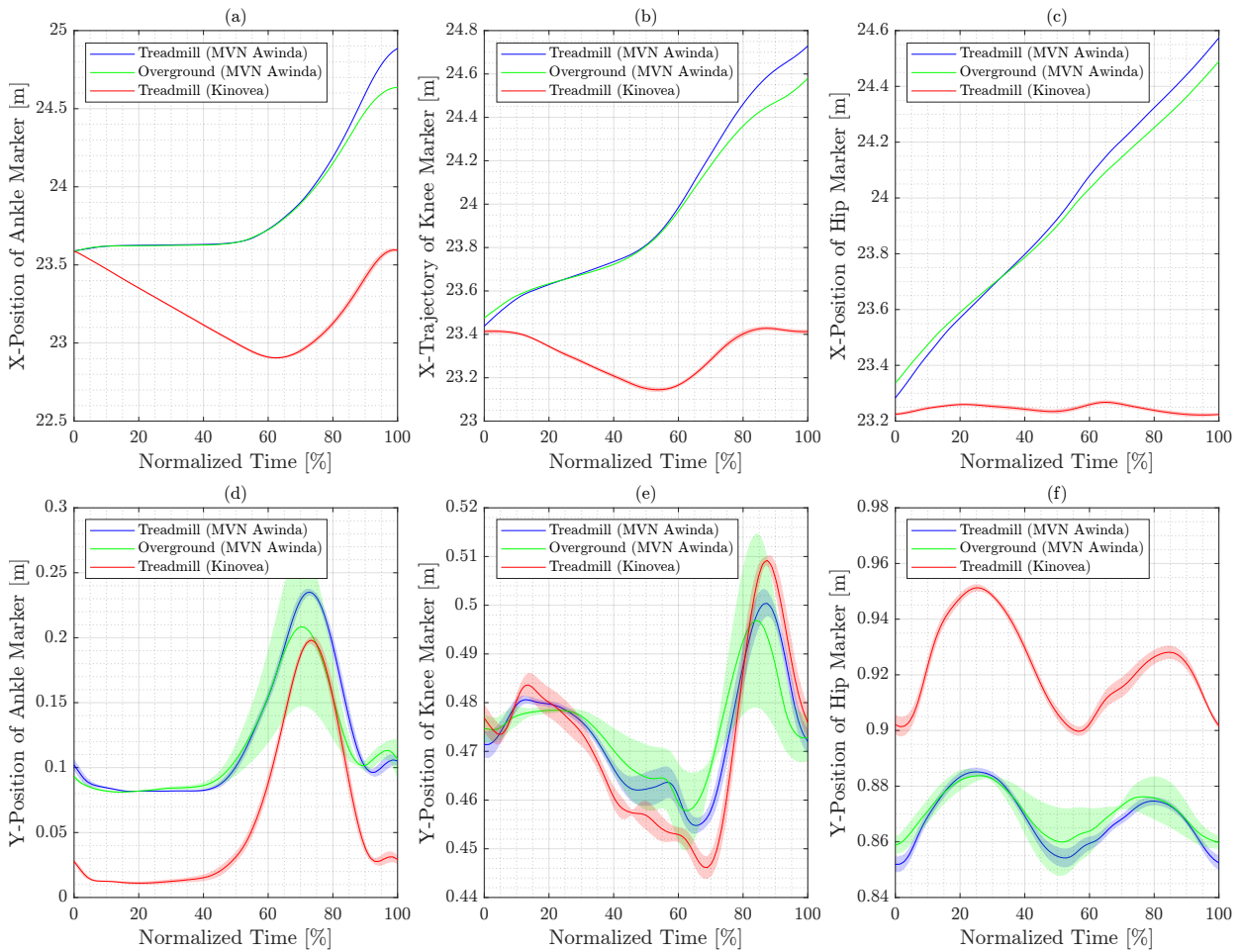


Fig. 4.15: Trajectories of the ankle, knee and hip markers in the horizontal (x) and vertical (y) axes throughout the GC.

The correctness of the GC segmentation was verified by analyzing the angular displacement of the knee joint. The kinematic characteristics of the knee joint throughout the GC has been extensively documented, so they can be used as a reference to determine whether the approach used for GC segmentation produced acceptable results. Initially, it was thought to use the trajectory of this joint to perform the segmentation; however, when analyzing the acquired data, it was observed that in some gait trials the shape of this trajectory

was significantly different, especially in a later stage of the study when abnormal gait was simulated, so it was decided to use the hip marker trajectory, which had a more consistent characteristic.

The angular displacement of the knee joint at each instant of time is determined by computing the orientation of the thigh and the shank and then calculating their difference. The orientation of a body segment is calculated by the following expression:

$$\theta_{si} = \begin{cases} \arctan\left(\frac{\Delta x}{\Delta y}\right), & \text{for } \Delta x \geq 0 \text{ and } \Delta y \geq 0 \\ \pi - \arctan\left(\frac{\Delta x}{\Delta y}\right), & \text{for } \Delta x < 0 \text{ and } \Delta y \geq 0 \\ \arctan\left(\frac{\Delta x}{\Delta y}\right) + \pi, & \text{for } \Delta x < 0 \text{ and } \Delta y < 0 \\ 2\pi - \arctan\left(\frac{\Delta x}{\Delta y}\right), & \text{for } \Delta x \geq 0 \text{ and } \Delta y < 0 \end{cases} \quad (4.10)$$

where θ_{si} is the angle of the segment i with respect to the horizontal axis, and Δx and Δy refer to the difference between the position of the markers located at the ends of the analyzed segment along the x and y axes, respectively.

Finally, the angular displacement of the knee joint is calculated by using the following expression:

$$\theta_{knee} = \theta_{shank} - \theta_{thigh} \quad (4.11)$$

Fig. 4.16 shows the angular displacement of the knee joint for each gait trial of the same individual. As with marker trajectories, the angle of the knee joint is averaged over a GC and its standard deviation calculated.

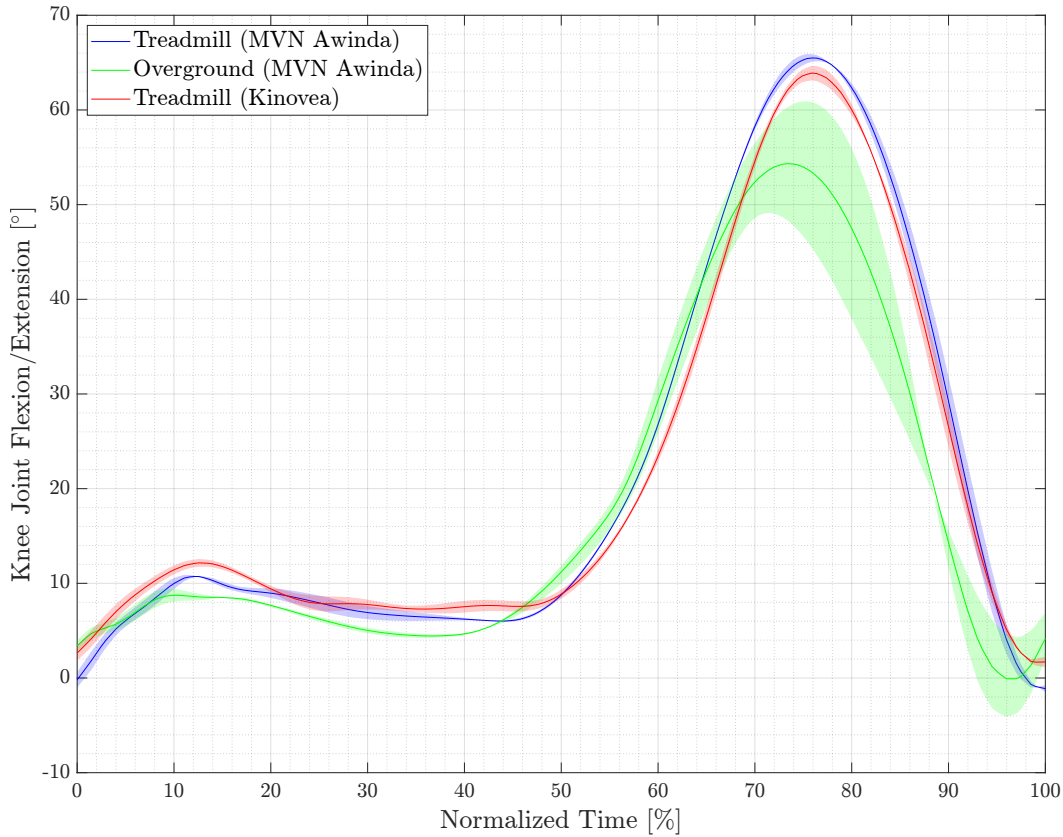


Fig. 4.16: Angular displacement of the knee joint throughout the GC.

In addition to the position of the hip, knee, and ankle markers, for the calculation of the ICR position of the thigh and leg at each instant of time, it is necessary to calculate the velocity of the markers. The velocities of the markers along the vertical and horizontal axes are calculated by numerical derivation. Fig. 4.17 presents the velocity of the markers throughout a GC.

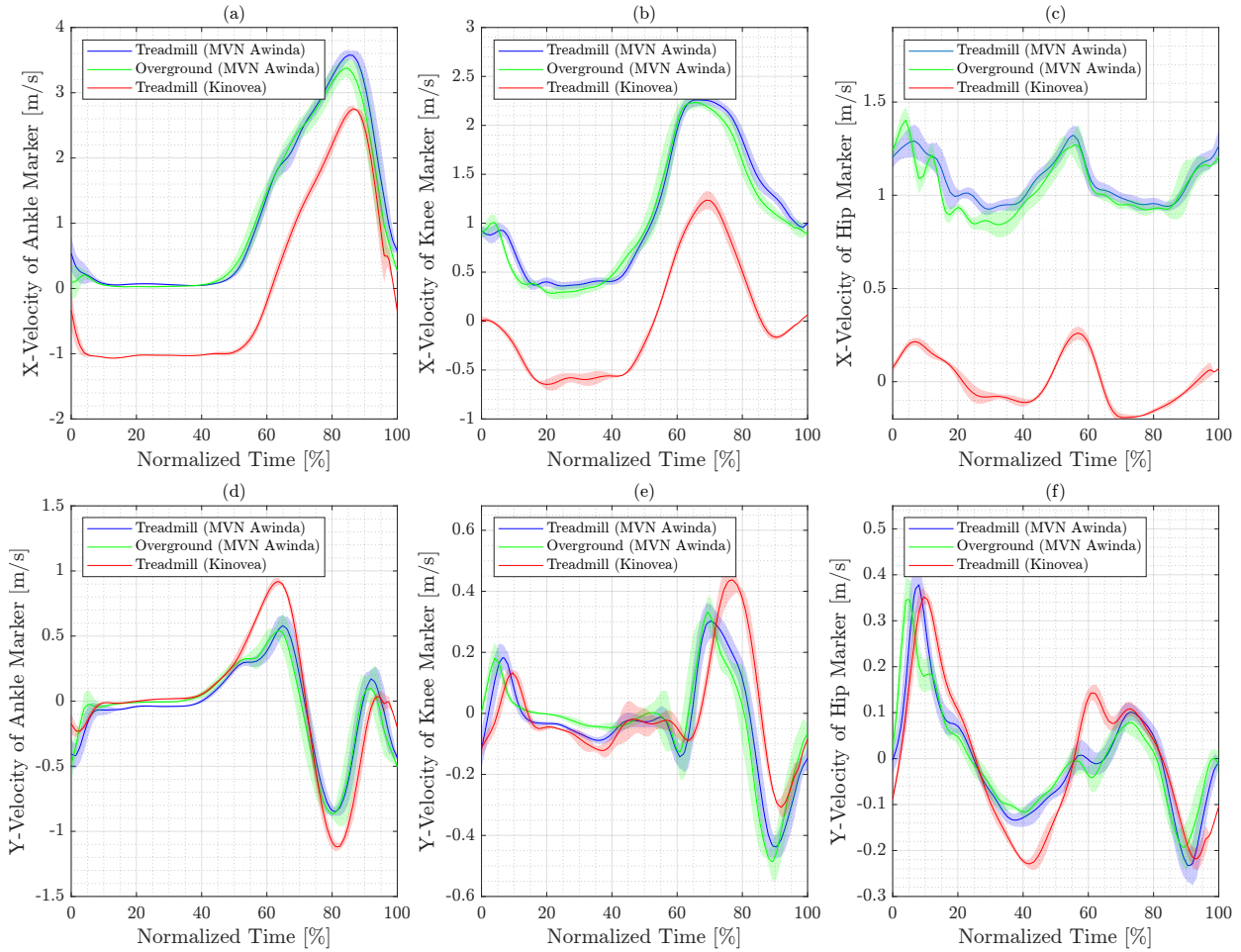


Fig. 4.17: Velocity of the ankle, knee, and hip markers in the horizontal (x) and vertical (y) axes throughout the GC.

As noted, a consistent pattern is observed in the three gait trials performed, and the data collected with the Kinovea tool differ the most in comparison to the others. Differences in marker position and velocity data when comparing data recorded using the MVN Awinda system and the Kinovea tool can be attributed in part to the placement of markers by the person conducting the gait trials. Although the markers were placed on the user according to a standardized protocol, their exact location is not error-free.

Furthermore, it is important to note that the position of markers linked to the MVN Awinda biomechanical model may not be directly aligned with markers tracked by Kinovea, i.e., the dimensions of the lower extremity in the MVN Awinda biomechanical model do not necessarily correspond to the relative distances between markers calculated within the video using Kinovea, which may even vary slightly from frame to frame. This discrepancy becomes apparent when examining the vertical trajectory of the hip marker, which exhibits a difference of about 5 cm (see Fig. 4.15).

4.4.1 Analysis of ICR trajectories in time domain

Since the representation of the position of a point in the plane requires the specification of two coordinates, ICR trajectories must be analyzed in terms of two features, which are time dependent. The choice of features used to specify the ICR position of the lower limb segments at a given time depends on the coordinate system used. In this study, we express the calculated ICR trajectories using Cartesian coordinates. This choice derives from the format of the acquired data, in which the position and velocity of anatomical markers are provided in Cartesian coordinates.

To fully define the position of a point in the plane, two coordinates must be specified. Since the ICR is a kinematic parameter related to the planar motion of a rigid body, each ICR trajectory can be decomposed into two time-dependent features. The choice of features used to specify the ICR position of the lower limb segments at a given time depends on the coordinate system used. In this work, the calculated ICR trajectories are initially expressed in Cartesian coordinates, as the position and velocity of the anatomical markers used for the calculation are provided as input to the algorithm in this format. In Cartesian coordinates, when using a conventional reference system, the axes used to describe the motion of a body are the vertical (y) and horizontal (x) axes.

Fig. 4.18 presents the ICR position of the shank and thigh determined for four selected time points. The position of the shank and thigh ICR at each time sample is determined by finding the intersection between two lines passing through the markers and perpendicular to the velocity vectors of the corresponding markers.

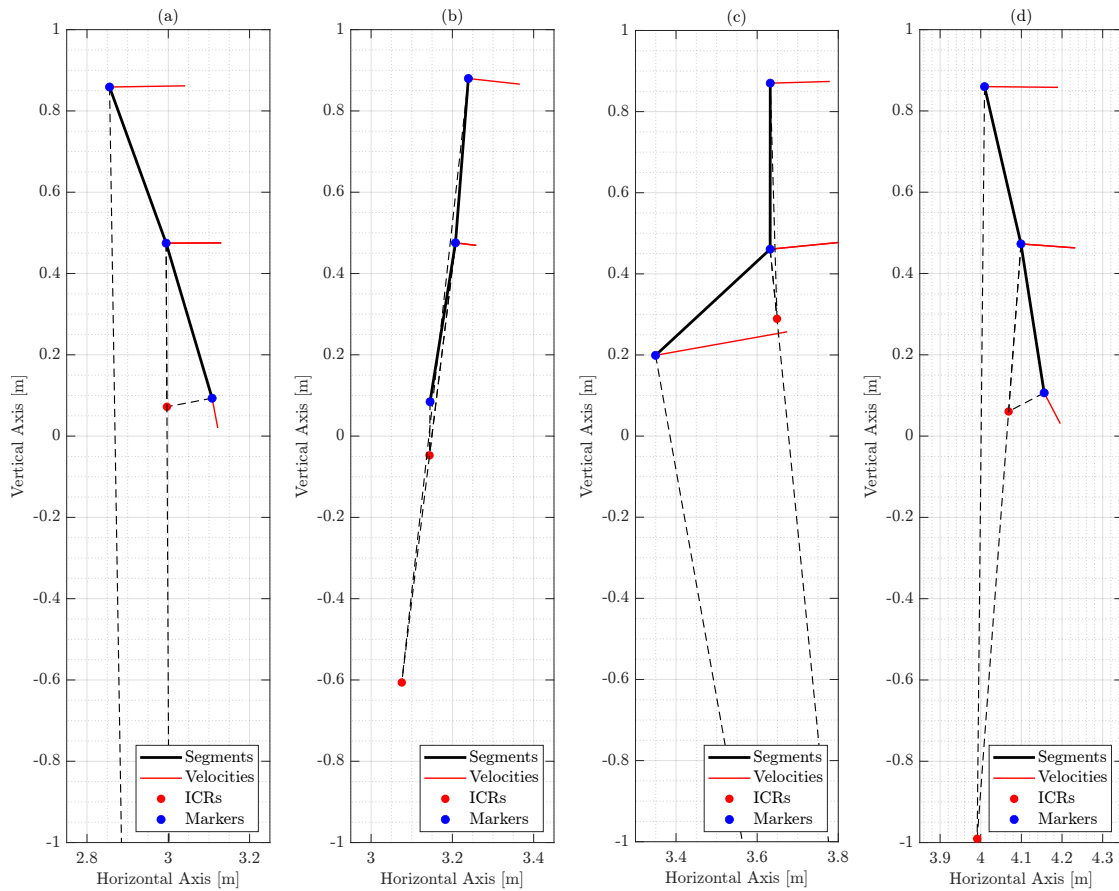


Fig. 4.18: ICR position of the shank and thigh around (a) 5%, (b) 35%, (c) 65% and (d) 95% of the GC

As can be seen in Fig.4.18, the position of the ICR of the lower limb segments (represented by the red dots) on the vertical and horizontal axes can deviate greatly from the position of the markers forming those segments. Very large deviations correspond to asymptotes that occur when the segment undergoes pure translation, which is observed in Figs. 4.18(a) and (c), in which the velocity vectors of the markers (represented by red lines) are nearly parallel.

Data collected using both the vision-based method and the inertial MoCap system are linked to a global reference frame, which is chosen by the person responsible for performing the gait trials. The absolute position of each marker is measured with respect to the origin of that reference frame. In the case of the MoCap system, the origin of the reference frame is determined by the place in the space where the subject wearing the system was at the time of calibration. In the case of the vision-based method, the origin can be any point in the video chosen by the researcher.

The origin of the reference frame may vary slightly from one experiment to another, especially if there are small errors in camera adjustment, in the case of using the vision-based method. To minimize the impact of the overall experimental setup on the measured data, it was decided to transform the global reference frame into a local reference frame, associated with each subject. Consequently, the origin of the reference frame was repositioned to coincide with the geometric center of the segment of the lower extremity analyzed of the subject under study. Therefore, the position of the new reference frame is determined by the following expressions:

$$X_{New} = X_{GR} - X_{COG} \quad (4.12)$$

$$Y_{New} = Y_{GR} - Y_{COG} \quad (4.13)$$

where x and y refer to the horizontal and vertical axes, the suffix GR denotes the global reference frame, the suffix COG denotes the geometric center of the analyzed segment, and the suffix New refers to the new local reference frame.

Identifying the distinctive (mean) characteristic of the ICR trajectories in the sagittal plane throughout the GC is essential in this study. The approach to carry out this task involved determining the ICR position for each time instant, segmenting the GCs in the collected data and averaging the segmented signals considering each GC as a single realization. However, a major limitation was identified in this context: it is impractical to calculate basic statistical parameters such as mean and standard deviation directly from these time series. This limitation arises because of the combined effect of the presence of asymptotes and potential time lag between time series, completely altering the expected characteristic of the time series. Since the sum of two values, where at least one of the values is/approaches infinity, also results in a value close to infinity, the shorter time lag between asymptotes produces an unexpected result. This issue was addressed by calculating the mean trajectory of the ICR and its standard deviation along the GC from the mean marker position and velocity data, i.e., the segmentation and calculation of the mean and standard deviation are performed on the position and velocity of the markers and not directly on the computed ICR trajectories along each segmented GC. When these steps are performed, three time series are obtained corresponding to the mean ICR trajectory ($\overline{ICR}$), the mean ICR trajectory plus one standard deviation ($\overline{ICR} + \sigma$), and the mean ICR trajectory minus one standard deviation ($\overline{ICR} - \sigma$).

Figs. 4.19 and 4.20 show the ICR trajectories of the shank and thigh during a gait cycle for a randomly chosen subject. In these figures, comparisons can be made (1) between characteristic trajectories $\overline{ICR}$ and $\overline{ICR} \pm \sigma$ during the same gait trial, and (2) between the characteristic ICR trajectories of each type of gait trial, all for the same subject. To make it easier for the reader to compare the typical characteristic of the ICR trajectory of the shank and thigh during the different types of gait trials, only the mean ICR trajectories ($\overline{ICR}$) are shown in Figs. 4.21 and 4.22 and are superimposed on the same plot.

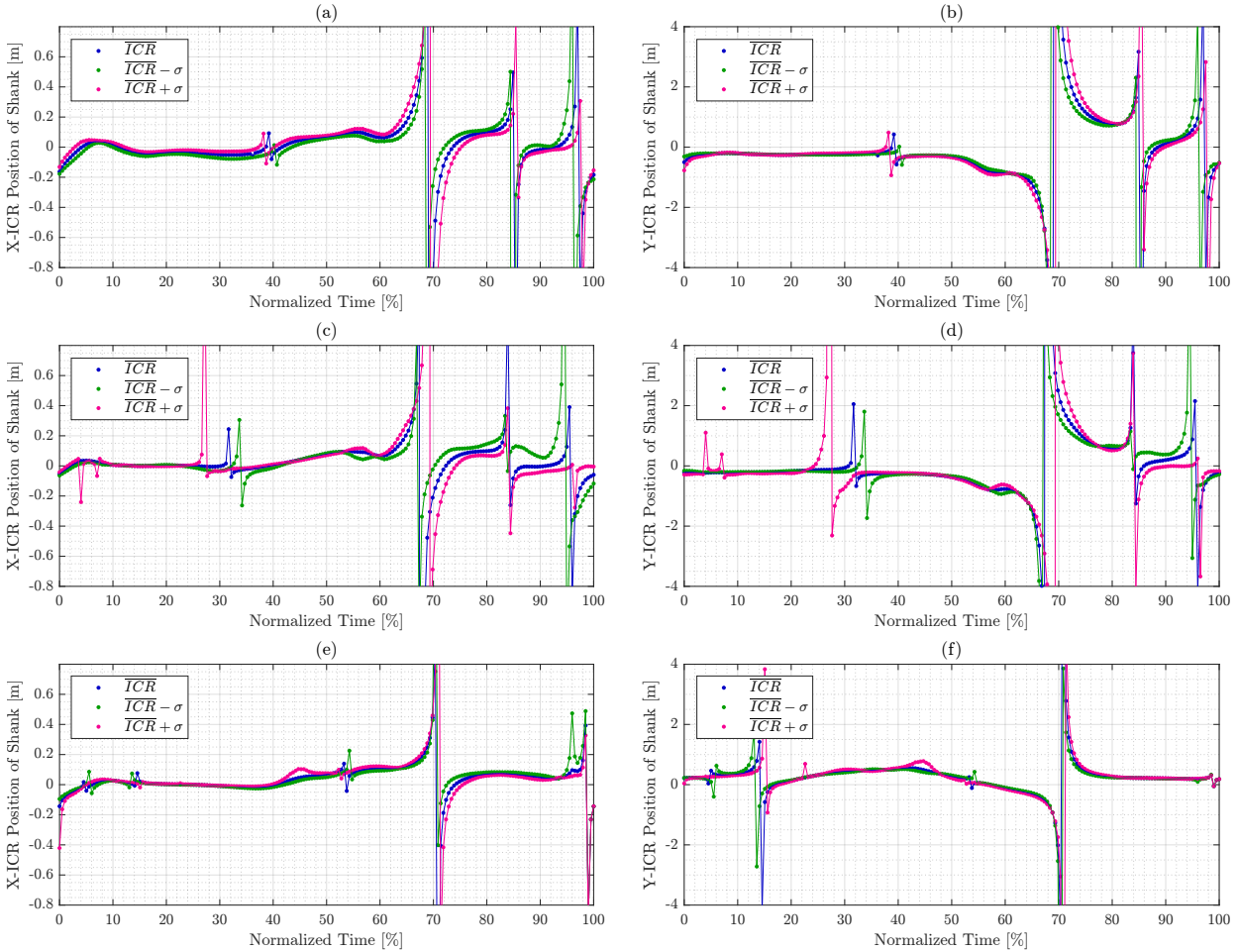


Fig. 4.19: Shank ICR trajectory during treadmill gait trials (MVN Awinda system) on the (a) horizontal (x) and (b) vertical (y) axes. Shank ICR trajectory during overground gait trials on the (c) horizontal (x) and (d) vertical (y) axes. Shank ICR trajectory during treadmill walking trials (Kinovea) in the (e) horizontal (x) and (f) vertical (y) axes.

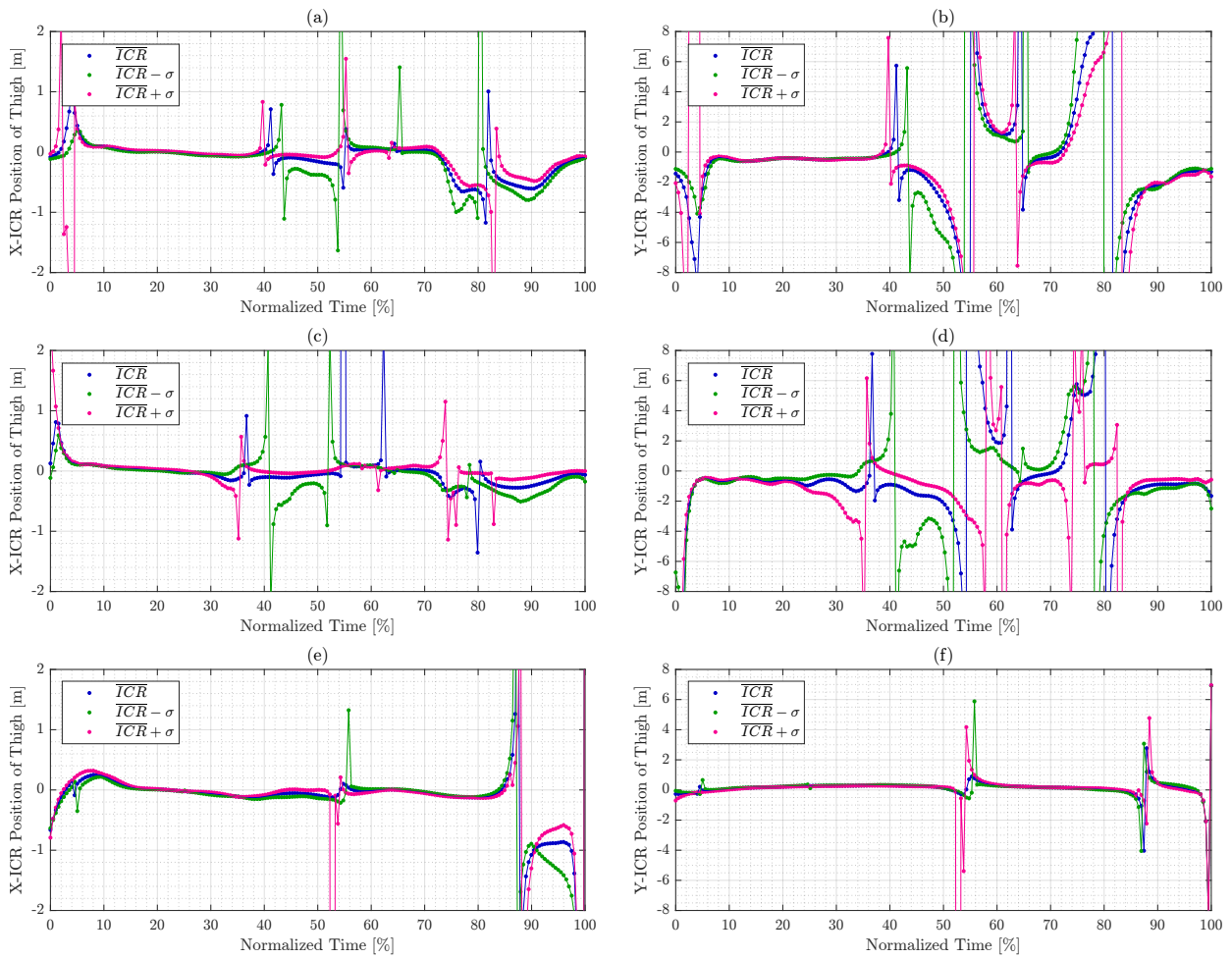


Fig. 4.20: Thigh ICR trajectory during treadmill gait trials (MVN Awinda system) on the (a) horizontal (x) and (b) vertical (y) axes. Thigh ICR trajectory during overground gait trials in the (c) horizontal (x) and (d) vertical (y) axes. Thigh ICR trajectory during treadmill walking trials (Kinovea) on the (e) horizontal (x) and (f) vertical (y) axes.

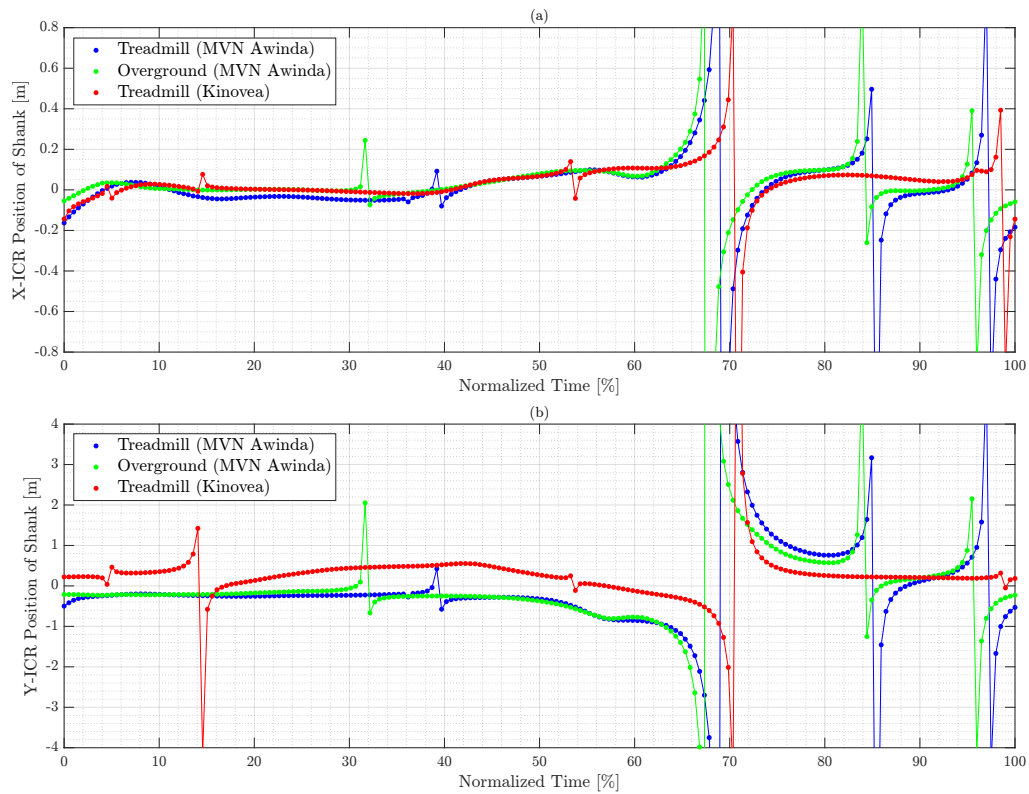


Fig. 4.21: Comparison of shank ICR trajectories in Cartesian coordinates during treadmill and overground gait trials using different data acquisition techniques. (a) Horizontal axis (x). (b) Vertical axis (y).

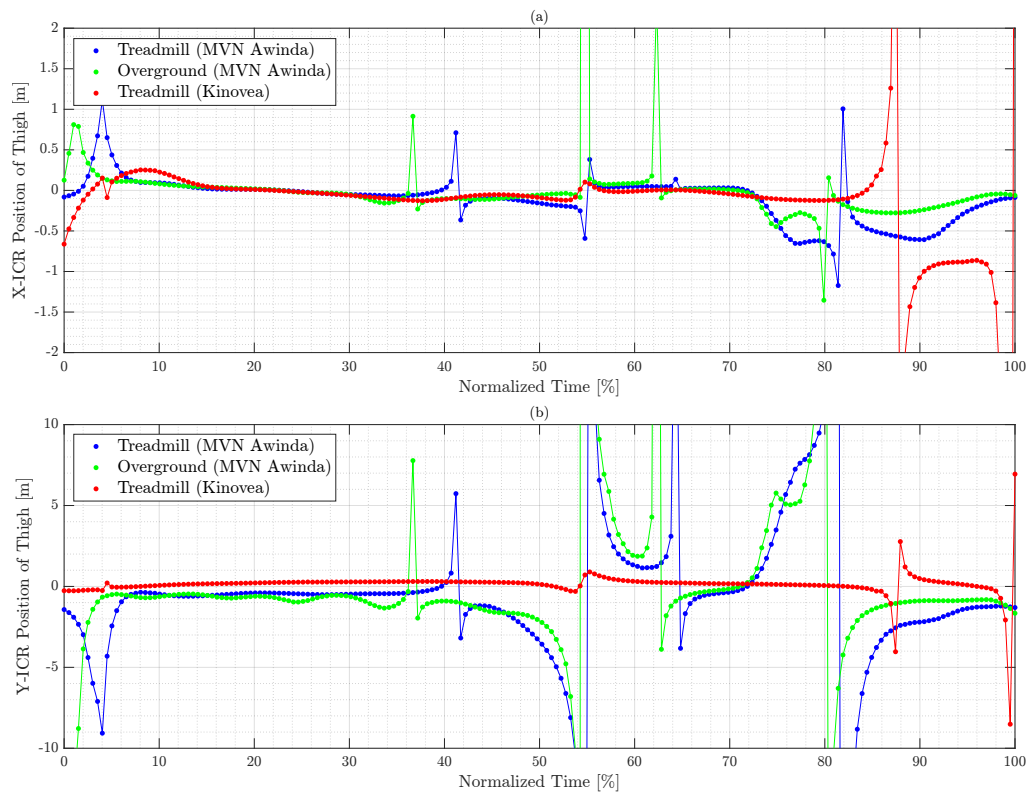


Fig. 4.22: Comparison of thigh ICR trajectories in Cartesian coordinates during treadmill and overground gait trials using different data acquisition techniques. (a) Horizontal axis (x). (b) Vertical axis (y).

The time series corresponding to the position of the shank ICR over the GC on both the horizontal and vertical axes show a similar characteristic when comparing data from different gait trials and when comparing the characteristic $\overline{ICR}$, $\overline{ICR} - \sigma$ and $\overline{ICR} + \sigma$ trajectories resulting from the same gait trial. Special attention should be paid to the time series obtained with the vision-based method. Unlike the data obtained with the other methods, it is observed that some asymptotes, such as the one occurring around 85% of the GC, do not occur. This is a curious fact, as more similarity is observed between the overground gait and the treadmill gait using the MVN Awinda system than between treadmill gait using the MVN Awinda system and treadmill gait using Kinovea. However, it is important to note that the MVN Awinda system operates with a principle that results in noticeable drift along the subject's forward axis in the collected data. Therefore, even when a subject walks on a treadmill that physically restricts forward motion, the data captured by the system depict the subject as moving forward without such constraints, similar to walking on the ground. In contrast, the vision-based method captures data that correspond more closely to natural observation, where the subject's forward motion is appropriately limited.

Figs. 4.20 and 4.22 show that, unlike the shank ICR trajectories, the thigh ICR trajectories on both axes contain larger discrepancies, especially when comparing data from different sources. The time series obtained with the Kinovea data appear smoother, and the signal amplitude is significantly lower compared to those obtained with the MVN Awinda system. From this visual inspection, we can state that a comparative analysis between data from the MVN Awinda system and the vision-based method used leads to erroneous inferences. Therefore, in subsequent analyzes, when the objective is to identify gait abnormalities, a quantitative analysis of the ICR trajectories is performed using only the data acquired from overground gait trials using the MVN Awinda system. The range of ICR position values presented in the plots has been limited to low-amplitude values, which characterize the signal most of the time. Due to the presence of asymptotes, the signal peaks reached values up to 20 times higher than the maximum values observed in the presented plots. What is important to rescue from the asymptotes found is their temporal location, since the amplitude of the data collected near the asymptote will be greater the closer it is to it, and the time series can take values of opposite sign or even 100 times greater despite the difference in the acquisition of the data being only one time sample.

The ICR position of a segment in the plane can also be defined using the polar coordinate system. This system is based on specifying the position of a point by the radius from a defined reference frame to the point and the angle with respect to a given orientation (typically the horizontal axis). The transformation from Cartesian to polar reference frame was performed to evaluate whether relevant information or information difficult to evidence in Cartesian coordinates can be extracted. The advantage of using polar coordinates is that the measured angle is restricted to values between 0° and 360° , which helps limit the amplitude of the time series. In contrast, Cartesian coordinates can result in an amplitude that extends from 0 to infinity, potentially complicating the analysis. In the context of 3D motion of rigid bodies, quaternions are frequently used to represent angular displacements because of their ability to handle the complexity of a 3D space without the ambiguity and discontinuity problems inherent in other angular representations. However, in this study, the analysis is limited to the sagittal plane, so the use of quaternions is not required. Instead, the use of polar coordinates with angular displacements expressed in terms of Euler angles is more appropriate.

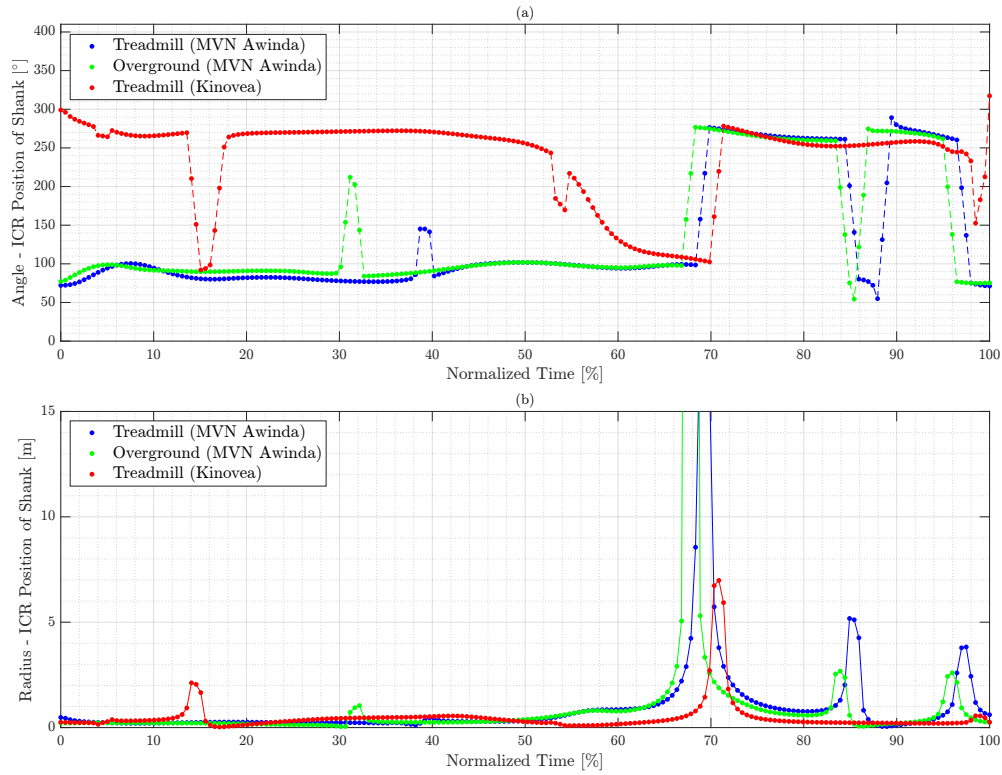


Fig. 4.23: Comparison of shank ICR trajectories in polar coordinates during treadmill and overground gait trials using different data acquisition techniques. (a) Angle. (b) Radius).

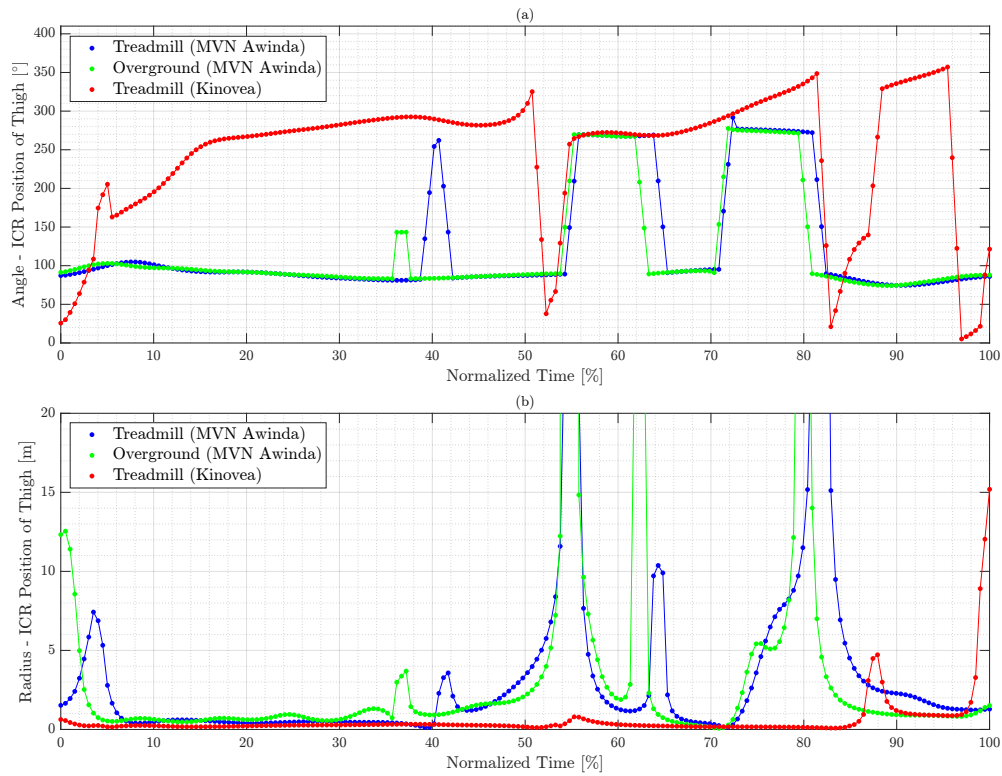


Fig. 4.24: Comparison of thigh ICR trajectories in polar coordinates during treadmill and overground gait trials using different data acquisition techniques. (a) Angle. (b) Radius.

Figs. 4.23 and 4.24 show the characteristic ICR trajectories in polar coordinates. As in the case of ICR trajectories expressed in Cartesian coordinates, in polar coordinates, the ICR trajectories show a similar characteristic when comparing the different data sources. Similarly, the ICR time series of the thigh show chaotic behavior, especially when attention is paid to the angle. Based on this fact, it can be inferred that possibly the exclusive use of the shank ICR trajectories when comparing gait trials between several subjects will lead to a more accurate analysis.

To this point, the ICR trajectories shown correspond to the gait-related data of the same subject, whose kinematic data were acquired by different approaches. However, the data obtained for several subjects should be compared to check whether the characteristic ICR trajectory is the same regardless of the subject evaluated if the gait trial is conducted under the same conditions. Figs. 4.25, 4.26, 4.27, and 4.28 show the ICR trajectories of the shank and thigh in Cartesian coordinates for each of the subjects, computed from data acquired using the MVN Awinda system. Figs. 4.25 and 4.26 show the time series corresponding to the ICR trajectories during treadmill gait trials, while Figs. 4.27 and 4.28 refer to the ICR trajectories during overground gait trials.

As expected, the characteristic patterns in the time series corresponding to the shank ICR trajectories are similar enough to base a comparative gait analysis of different subjects on this type of data. Although time lags are observed, the average characteristic is easily observable. However, atypical time series (outliers) are observed as in the case of subject 4 which seems to present an asymptote around 10% of the GC, which is not present in the rest of the time series (see Fig. 4.25). On the other hand, the time series related to the thigh ICR are completely chaotic. When comparing, very uneven amplitudes are observed and even the time lags are considerable high, making it difficult to extract a representative mean characteristic.

Fig. 4.27 also shows atypical time series. In this figure, subject 10 appears to present an asymptote around 18% of the GC. However, despite this fact, the shank ICR trajectories during the overground gait have a characteristic average pattern that can be observed simply by visual inspection. As in the case of treadmill gait trials, the time series corresponding to the thigh ICR trajectories during overground gait trials, shown in Fig. 4.26, are chaotic and it is difficult to rescue a pattern in the observed time series.

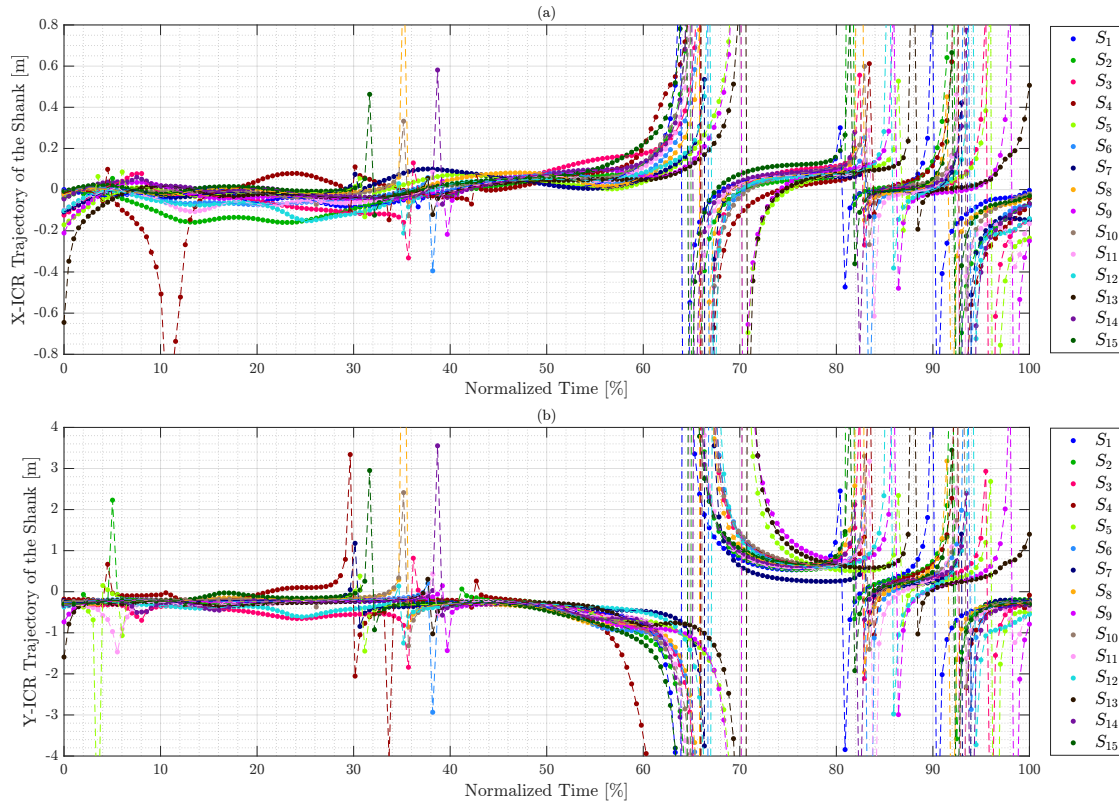


Fig. 4.25: Shank ICR trajectories in Cartesian coordinates of each subject during treadmill gait trials. (a) Horizontal (x) axis. (b) Vertical (y) axis.

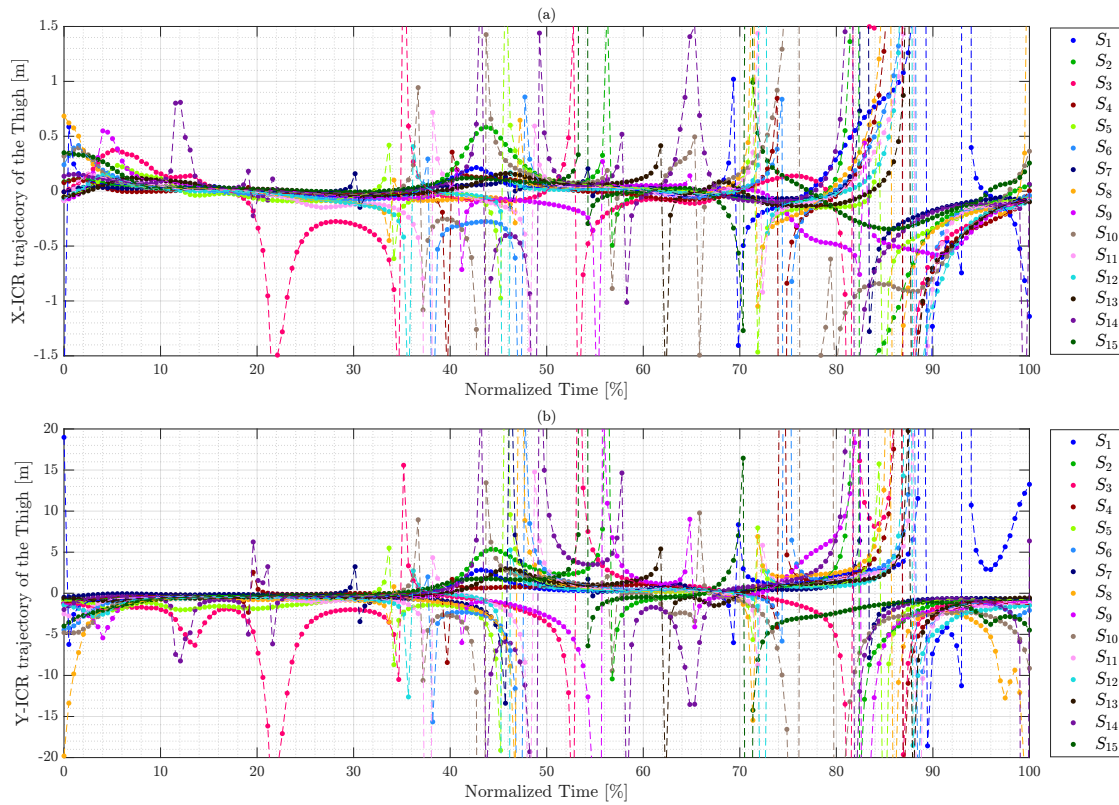


Fig. 4.26: Thigh ICR trajectories in Cartesian coordinates of each subject during treadmill gait trials. (a) Horizontal (x) axis. (b) Vertical (y) axis.

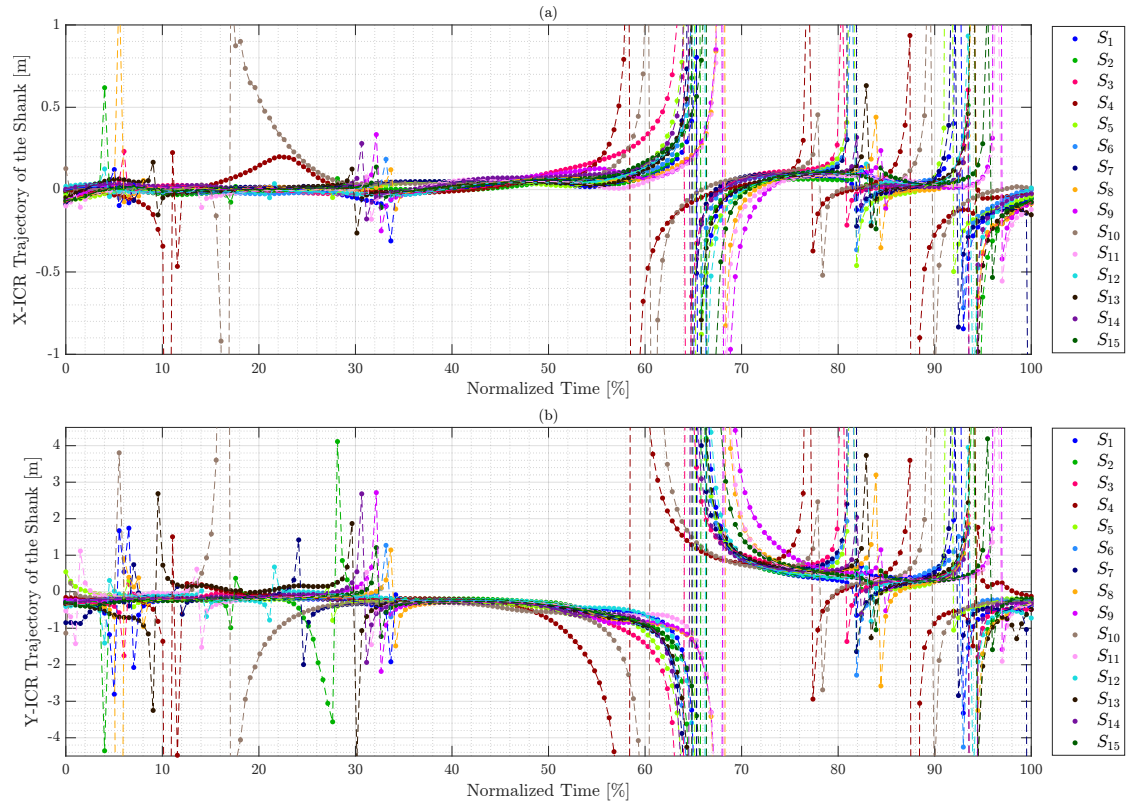


Fig. 4.27: Shank ICR trajectories in Cartesian coordinates of each subject during overground gait trials. (a) Horizontal (x) axis. (b) Vertical (y) axis.

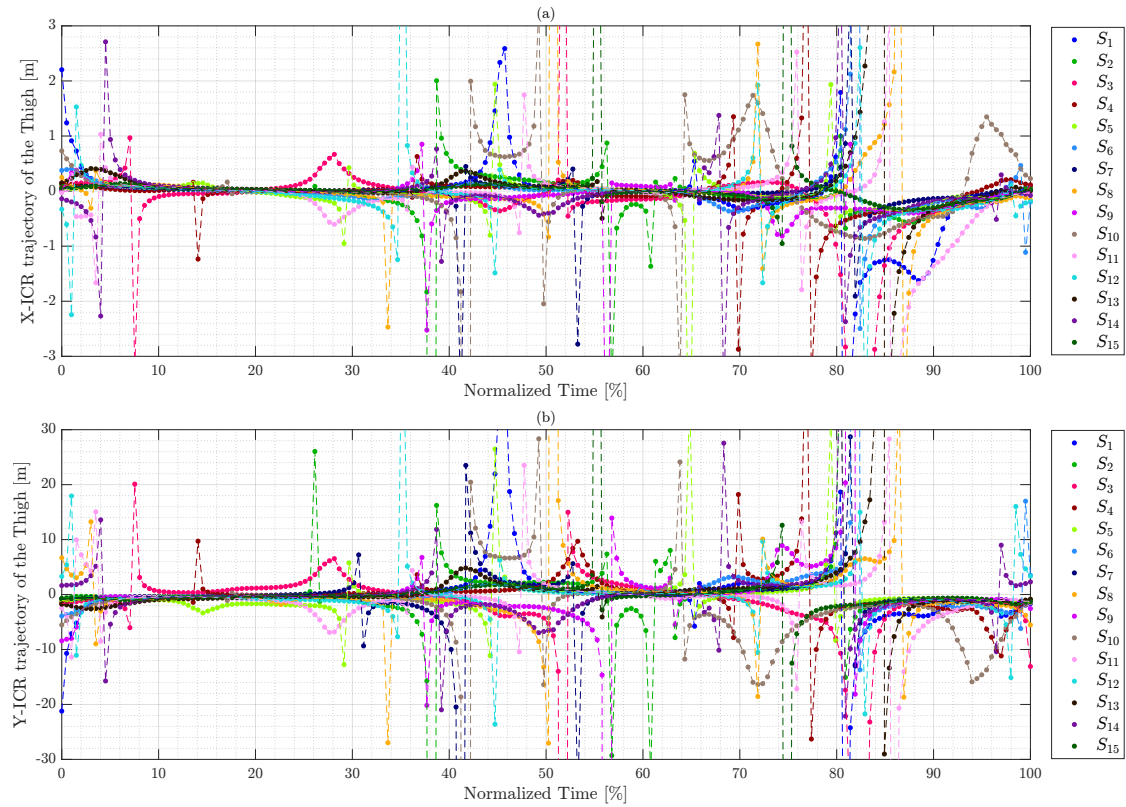


Fig. 4.28: Thigh ICR trajectories in Cartesian coordinates of each subject during overground gait trials. (a) Horizontal (x) axis. (b) Vertical (y) axis

Using the same approach as that used to segment the trajectories of the anatomical markers of the lower extremities (hip, knee, and ankle) in their respective GCs, the segmentation of the subphases contained in the time series corresponding to the ICR trajectories of the shank is carried out. On this occasion, the task consists of finding the temporal location at which the asymptotes occur and dividing the time series into fragments found between two asymptotes. Figs. 4.29, 4.30 and 4.31 show the segmented subphases of the shank ICR trajectories for the same subject in Cartesian coordinates. The plots presented correspond to the time series computed from the data collected during the treadmill gait trial using the MVN Awinda system, the overground gait trial, and the treadmill gait trial using Kinovea, respectively.

Each figure is made up of several subplots, where the upper subplots correspond to the trajectory of the shank ICR on the horizontal axis, and the lower subplots correspond to the trajectory of the shank ICR on the vertical axis. From left to right, we observe, in order of occurrence in the GC, each of the subphases obtained. The advantage of segmenting the shank ICR trajectories into their subphases is that it is possible to compare two trajectories by comparing their subphases and thus obtain a more accurate measure of similarity between time series.

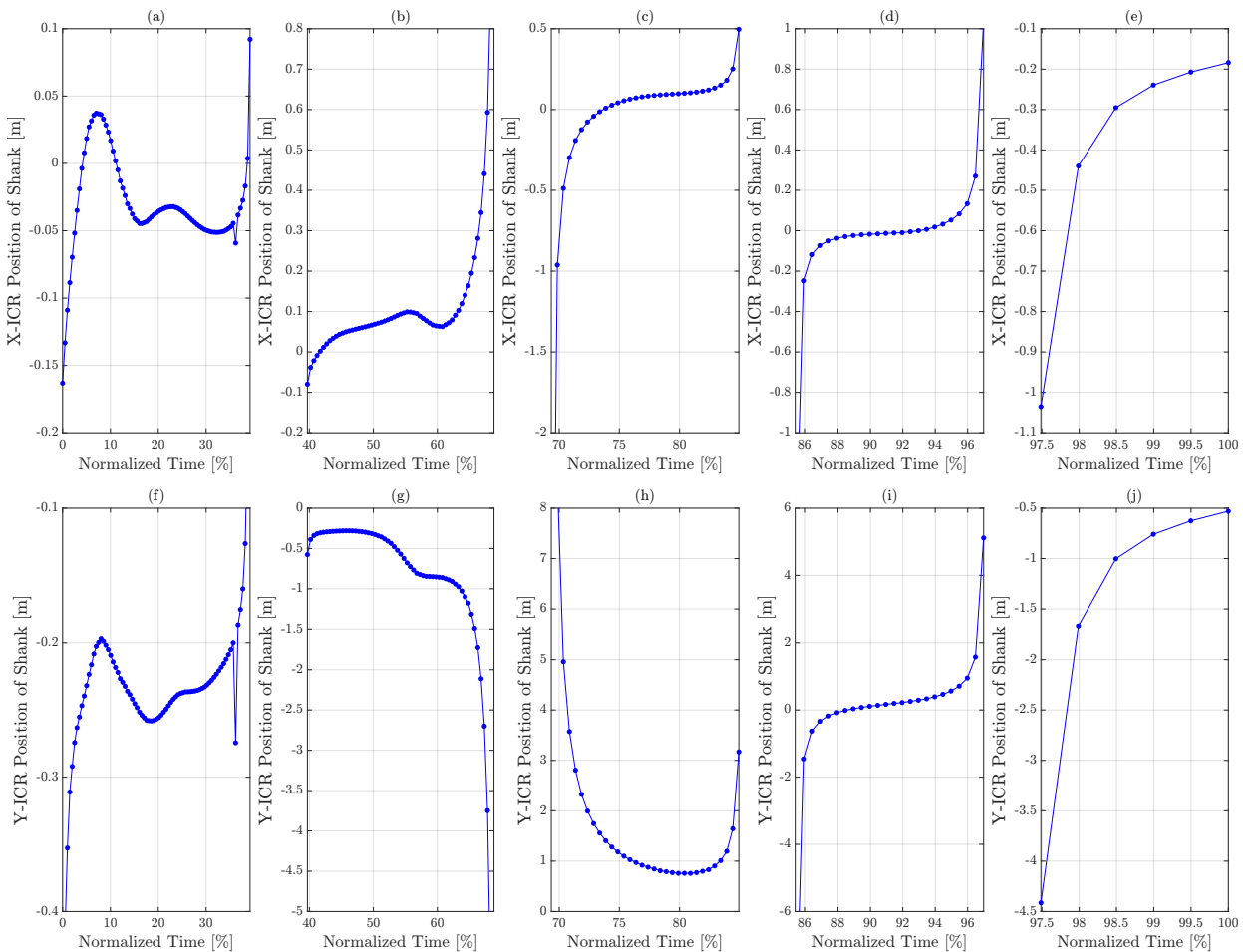


Fig. 4.29: Segmented gait intervals obtained from data collected during treadmill gait trials using the MVN Awinda system. At the top, the shank ICR trajectory is shown on the x axis during the interval (a) 1st, (b) 2nd, (c) 3rd, (d) 4th and (e) 5th. At the bottom, the shank ICR trajectory is shown on the y axis during the interval (f) 1st, (g) 2nd, (h) 3rd, (i) 4th and (j) 5th.

In Figs. 4.29(e) and 4.30(e), it can be noticed that, in addition to showing a similar pattern between gait trials, the segmented interval is quite short. This interval should not be considered as an additional phase of the GC, since it does not lie between two asymptotes, so the junction between subplots (a) and (e), and (f) and (j) is treated as one phase in the subsequent discussion. Due to the junction between these two intervals, the start of the GC is the end of the interval shown in Figs. 4.29(d) and 4.30(d). The first segmented interval, corresponding to the junction of plots (a) and (e), ranges approximately from 0 to 30% of the GC, which comprises the combination of loading response and mid-stance, studied in Chapter 2. The next interval ranges from 30% to around 65% of the GC and includes the terminal stance and pre-swing phases up to slightly more than the early swing phase (see Figs. 4.29(b) and 4.30(b)). The next interval ranges from 65% to 85% and corresponds to the combination of early swing and mid-swing (see Figs. 4.29(c) and 4.30(c)). Finally, the last interval corresponds to the terminal swing (see Figs. 4.29(d) and 4.30(d)).

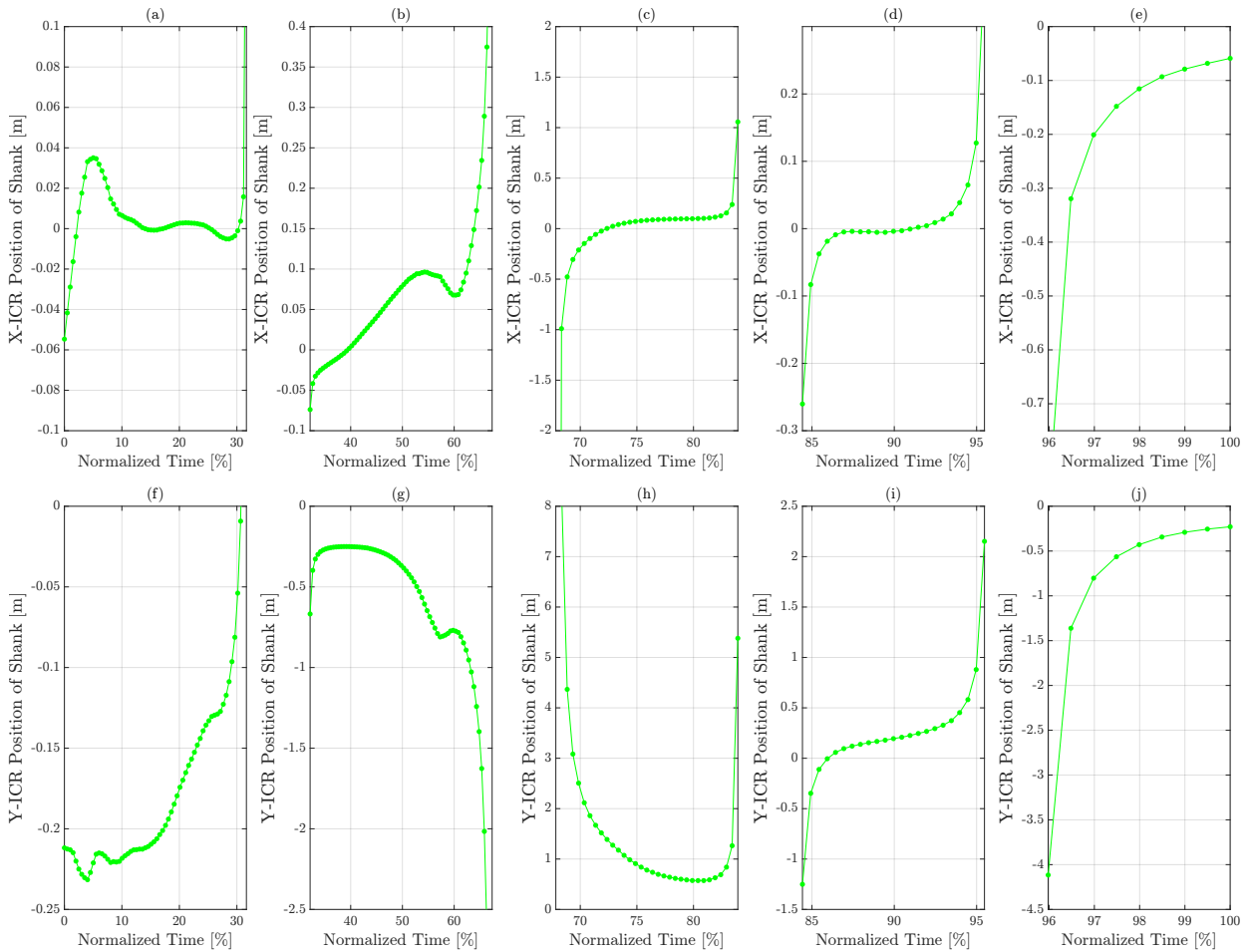


Fig. 4.30: Segmented gait intervals obtained from data collected during overground gait trials using the MVN Awinda system. At the top, the shank ICR trajectory is shown on the x axis during the interval (a) 1st, (b) 2nd, (c) 3rd, (d) 4th and (e) 5th. At the bottom, the shank ICR trajectory is shown on the y axis during the interval (f) 1st, (g) 2nd, (h) 3rd, (i) 4th and (j) 5th.

The segmented intervals resulting from the analysis of ICR trajectories computed from treadmill and overground gait trials using the MVN Awinda system have practically the same characteristics, with only the

interval length varying approximately 2%. Similarly, the characteristics of the segmented intervals computed from the data collected using Kinovea are considerably different.

In Fig. 4.31 it can be seen that the last interval (see Fig. 4.31(e)) and the first interval (see Fig. 4.31(a)) can be combined as was done with intervals (a) and (e) in the case of the data acquired from the gait trials in which the MVN Awinda system was used. The first interval resulting from the segmentation of shank ICR trajectories ranges from 0% to about 15% of the GC which corresponds to the loading response and a portion of mid-stance (Junction of Fig. 4.31(a) and Fig. 4.31(e)). The next interval ranges from 18% to 50% and corresponds to the next portion of the mid-stance along with the terminal-stance (see Fig. Fig. 4.31(b)). The next interval corresponds to the junction of pre-swing and initial swing (see Fig. 4.31(c)), while the last interval corresponds to the mid-swing and terminal swing (see Fig. Fig. 4.31(d)).

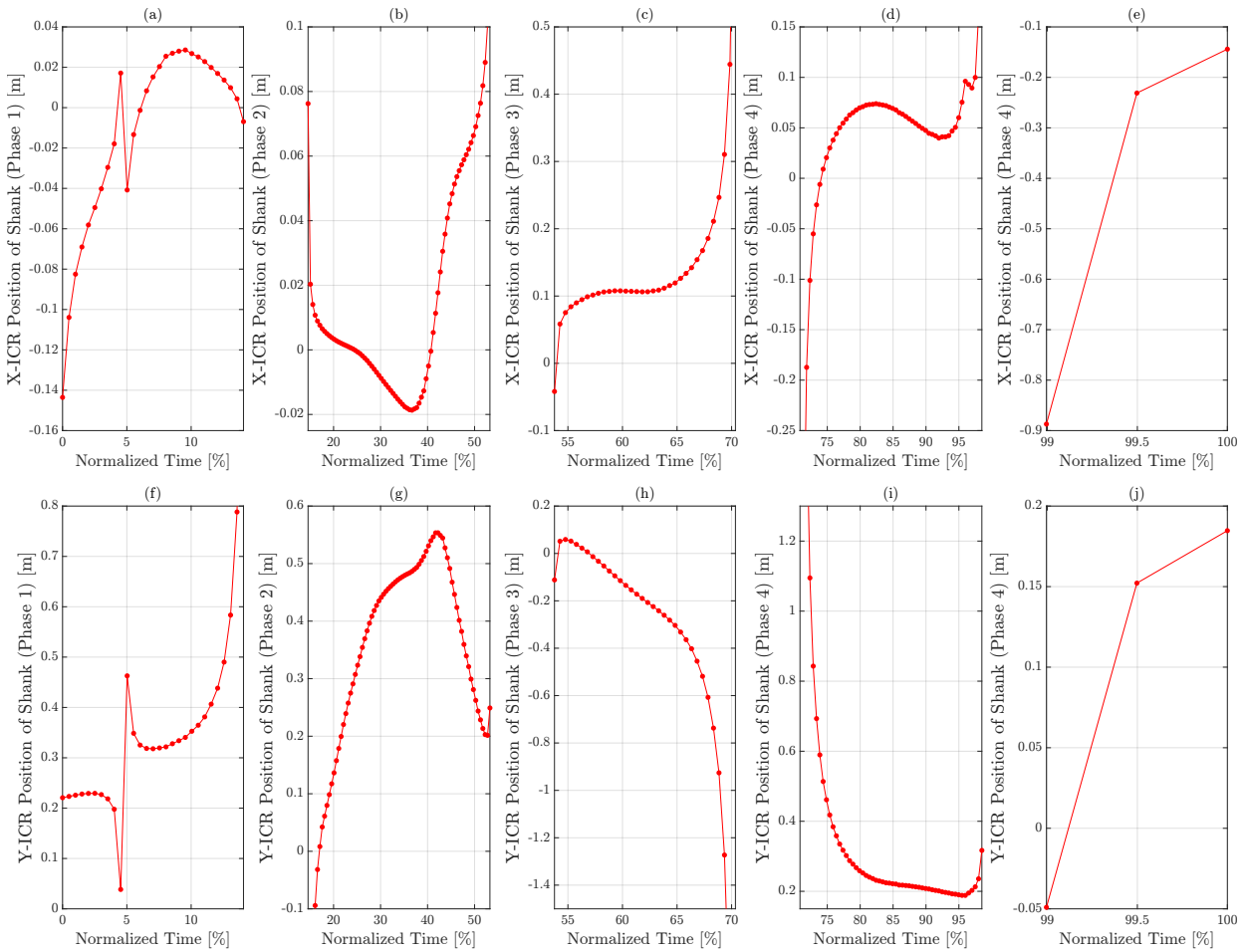


Fig. 4.31: Segmented gait intervals obtained from data collected during treadmill gait trials using Kinovea. At the top, the shank ICR trajectory is shown on the x axis during the interval (a) 1st, (b) 2nd, (c) 3rd, (d) 4th and (e) 5th. At the bottom, the shank ICR trajectory is shown on the y axis during the interval (f) 1st, (g) 2nd, (h) 3rd, (i) 4th and (j) 5th.

From the analysis of the segmented intervals, it is clear that there is a relationship between the occurrence of major events throughout the GC and the occurrence of asymptotes in the shank ICR trajectories in the sagittal plane. This fact is especially evident when comparing the gait phase intervals and the segmented intervals from the overground gait trials.

The trajectories of the ICR of the shank and thigh in the sagittal plane have, up to this point, been analyzed by independent examination of their horizontal and vertical components. However, these trajectories constitute a planar motion of the ICR in the plane when viewed as a whole. Fig. 4.32 presents the trajectory of ICR of the shank and thigh in the sagittal plane.

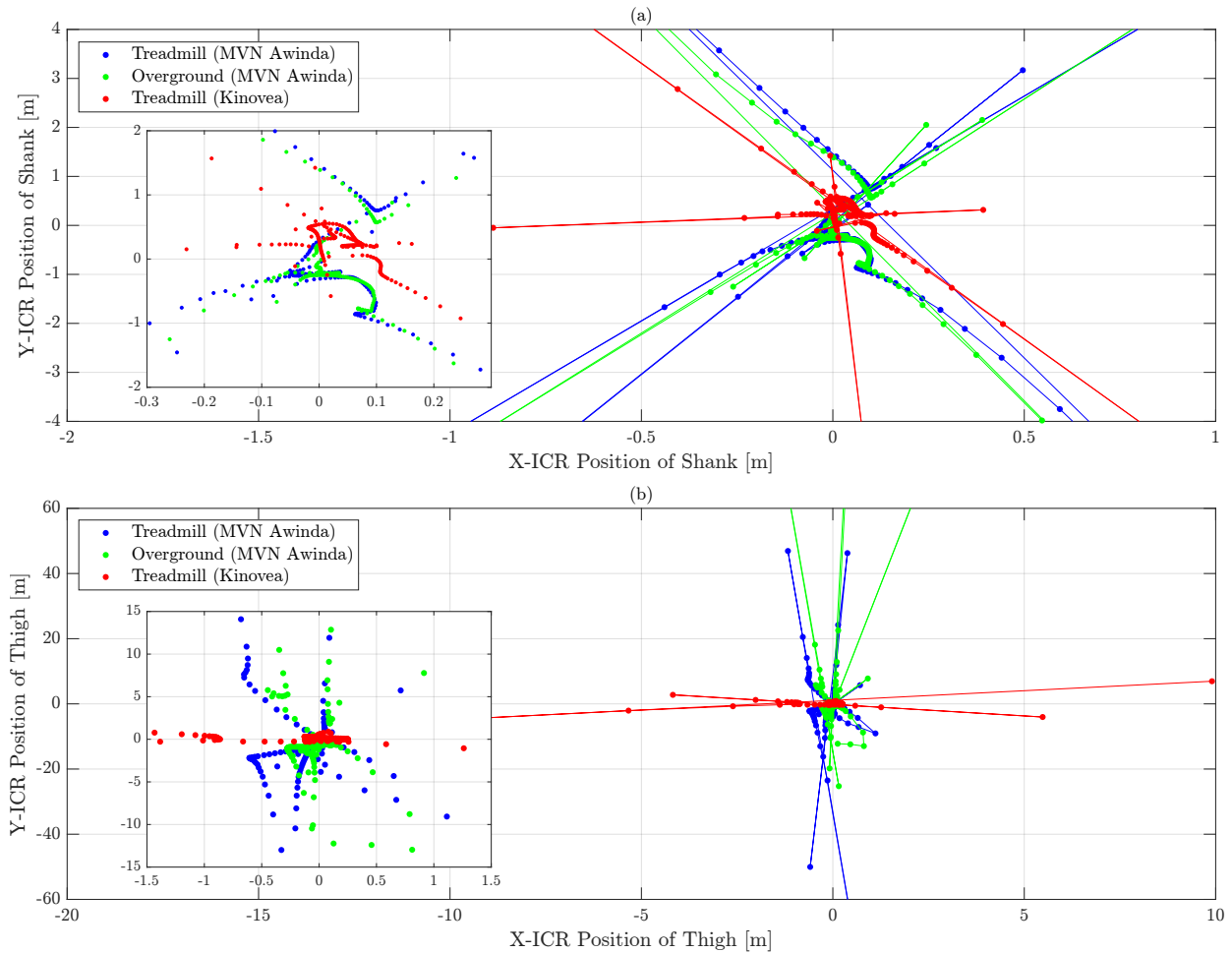


Fig. 4.32: Planar ICR trajectory of (a) shank and (b) thigh during different gait trials.

As can be seen, the presence of asymptotes makes it difficult to examine large-scale trajectories. Therefore, each figure presents a smaller-scale subplot in which the ICR trajectories during each gait trial are observed. The ICR trajectories of the shank present the same pattern when comparing the gait trials in which the MVN Awinda system was used, while the shank ICR trajectories computed from the data recorded with Kinovea are particularly different. In the case of the thigh ICR trajectories, the patterns do not coincide and again show chaotic behavior.

As preliminary conclusions up to this point, the ICR trajectories of the thigh may not be helpful in discriminating altered gait patterns, as it is difficult to recognize a reference pattern.

4.4.2 Frequency content of ICR trajectories

Any time series can be decomposed into simple harmonic waves with different frequencies, amplitudes, and phases. While periodic signals can be decomposed into a series of simple sine waves with discrete frequencies, nonperiodic signals can be decomposed into a combination of an infinite number of simple sine waves with continuous real frequencies. This is the basic principle of the Fourier Transform (FT), which is a tool to transform a time-sampled signal into the same time-frequency sampled signal and examine its composition at each frequency. The FT is defined by the following expression:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \quad (4.14)$$

where $x(t)$ is the signal in the time domain, $X(f)$ is the transformed signal in the frequency domain, and f are the frequencies of the sine waves contained in the signal.

The FT is a very useful tool for signal analysis. However, one of the requirements for its use is that the analyzed signal be stationary, that is, that its statistics do not vary or vary slightly over time. Walking is an activity consisting of repetitive movements. Each characteristic pattern of a GC is repeated every so often, with slight variations from one step to another. Over a short period of time, the deviations in the gait patterns are small enough to consider the gait to be a stationary process. However, it is important to clarify that the characteristics associated with an individual's gait may change over time as a function of external factors, such as muscle fatigue.

In the analysis of nonstationary signals in the frequency domain, it is necessary to decompose the signal into simpler signals with different frequencies that also vary in time. This task can be performed using different approaches, such as Wavelet Analysis (WA) and Short Time Fourier Transform (STFT). The WA consists of decomposing the signal into wavelets, which are waves with different amplitudes, scales, and time lags (Torrence and Compo, 1998). The STFT, on the other hand, decomposes the signal into short time segments and calculates the conventional FT of each segment (Krishnan, 2021). Both techniques provide a dynamic perspective of the signal behavior, useful for analyzing nonstationary signals.

In this study, STFT was selected because of its suitability for analyzing periodic and relatively stable frequency components within short time windows. Gait signals typically exhibit consistent frequency patterns over short periods, making the ability of STFT to provide clear, localized frequency information particularly beneficial. WA, while providing adaptive resolution and capturing variable frequency content more accurately, introduces complexity that may not be necessary for gait analysis.

The general assumption in the implementation of STFT is that the signal segments to which FT is applied are stationary, which means that the frequency content does not vary over that interval. In STFT calculation, a sliding window $w(t)$ with M samples is applied to the signal to obtain small segments that are then transformed into the frequency domain using FT (See. Fig. 4.33). The mathematical definition of continuous STFT is given by the expression:

$$X_m(t, f) = \int x(\tau) w(\tau - t) e^{-j2\pi f\tau} d\tau \quad (4.15)$$

where $x(t)$ is the input signal at time t , $w(t)$ is the chosen window function, $X_m(t, f)$ is the calculated STFT of the windowed data centered on time t .

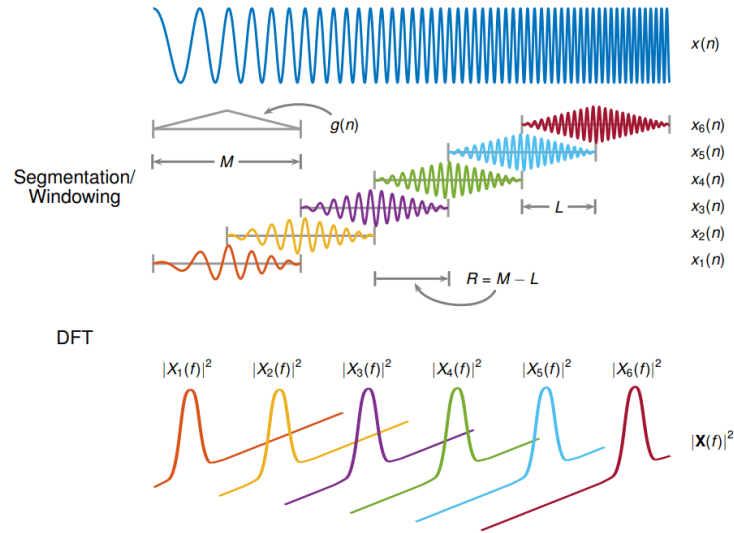


Fig. 4.33: Principle of Short-Time Fourier Transform (The MathWorks Inc., 2019).

Since in digital signal processing, the input signal is discretized in time, the implemented STFT algorithm also must be discretized. The mathematical definition of discrete STFT is given by:

$$X_m(k, m) = \sum_{n=0}^{N-1} w(n) x(n + mR) e^{-j\frac{2\pi}{N}kn} \quad (4.16)$$

where $x(n)$ is the input signal in sample n , $w(n)$ is the chosen window function, k is the discrete frequency, $X_m(k, m)$ is the STFT of the windowed data centered on sample mR , and R is the hop size, in samples between successive DTFTs.

The square of the magnitude of the FT is known as the Power Spectral Density (PSD), which describes how the energy of a signal is distributed across different frequencies. To visualize the PSD of a signal, it is typically plotted in a two-dimensional diagram where the x axis represents the frequency and the y axis represents the PSD values. In contrast, the STFT yields a power spectrum that varies with both frequency and time. This is visualized using a spectrogram, a two-dimensional plot where the x axis represents time, the y axis represents frequency, and the color or intensity of each point indicates the power at that specific frequency and time.

The choice of window length for the STFT is crucial, as it affects both temporal and frequency resolution. This parameter will basically determine the number of temporal and frequency bins shown in the spectrogram. A larger window provides better frequency resolution but poorer time resolution, while a shorter window improves time resolution at the cost of frequency accuracy. In this study, the frequency content of the calculated ICR trajectories of the shank and thigh are analyzed by using FT and STFT, considering stationary and nonstationary processes, respectively. In both cases, the frequency content was calculated and presented graphically. A Hamming window with a length of 10 samples and an overlap of 5 samples was used to calculate the STFT, while the DFT was calculated after zero-padding the signals until 256 samples were reached. To mitigate the impact of asymptotes on the calculated PSD, the magnitude of the signals was truncated to fall within the range of $[-a, a]$. The value of a was determined by visually inspecting the ICR trajectories in the time domain extracted from gait trials collected using the MVN Awinda system.

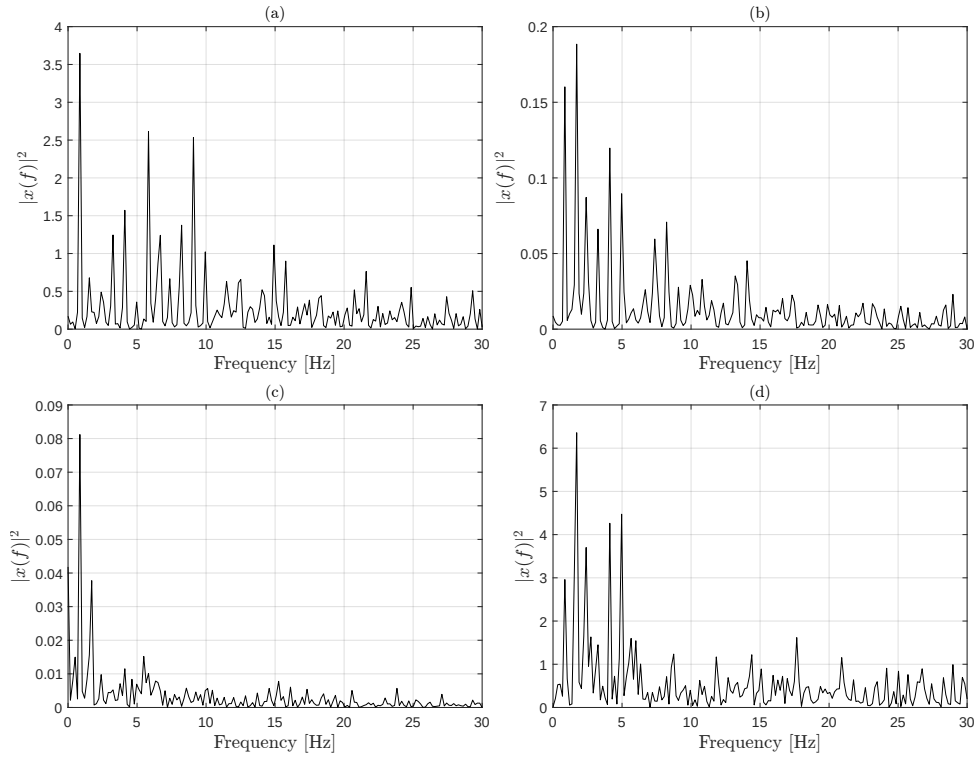


Fig. 4.34: PSD of ICR trajectories during treadmill gait trials collected using the MVN Awinda system. (a) Shank ICR trajectory on the horizontal axis. (b) Shank ICR trajectory on the vertical axis. (c) Thigh ICR trajectory on the horizontal axis. (d) Thigh ICR on the vertical axis.

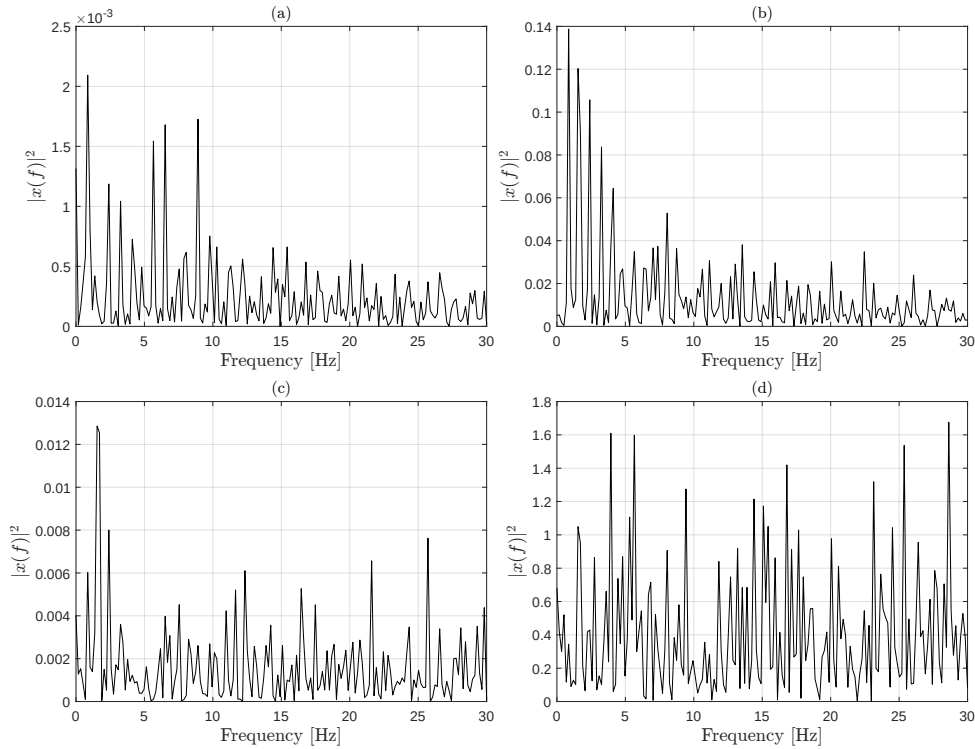


Fig. 4.35: PSD of ICR trajectories during overground gait trials collected using the MVN Awinda system. (a) Shank ICR trajectory on the horizontal axis. (b) Shank ICR trajectory on the vertical axis. (c) Thigh ICR trajectory on the horizontal axis. (d) Thigh ICR on the vertical axis.

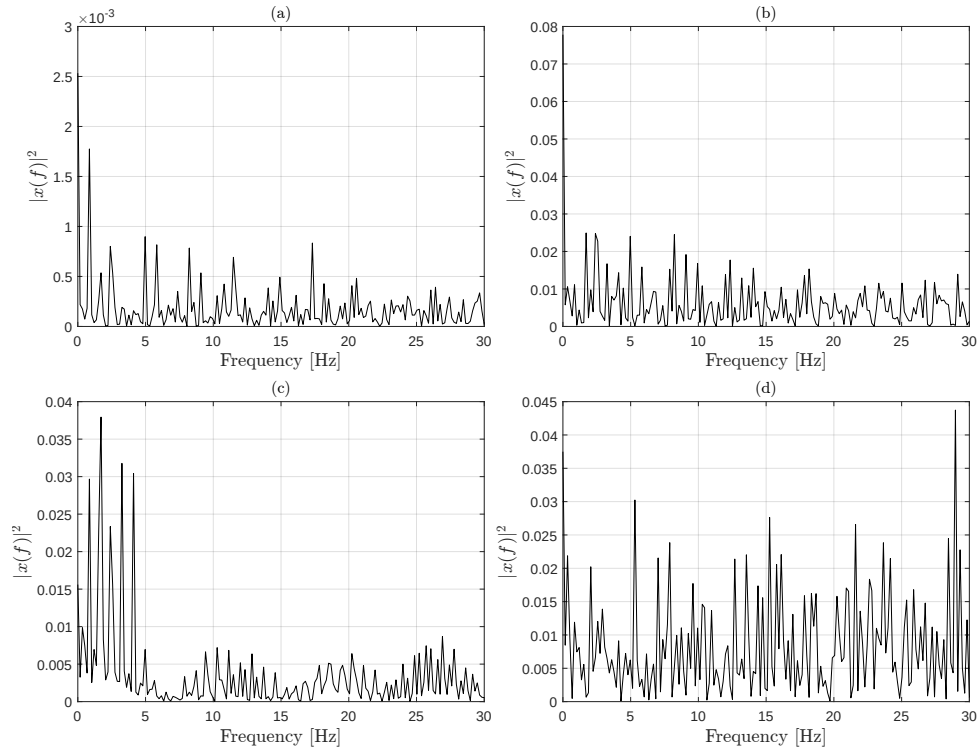


Fig. 4.36: PSD of ICR trajectories during treadmill gait trials collected using Kinovea. (a) Shank ICR trajectory on the horizontal axis. (b) Shank ICR trajectory on the vertical axis. (c) Thigh ICR trajectory on the horizontal axis. (d) Thigh ICR trajectory on the vertical axis.

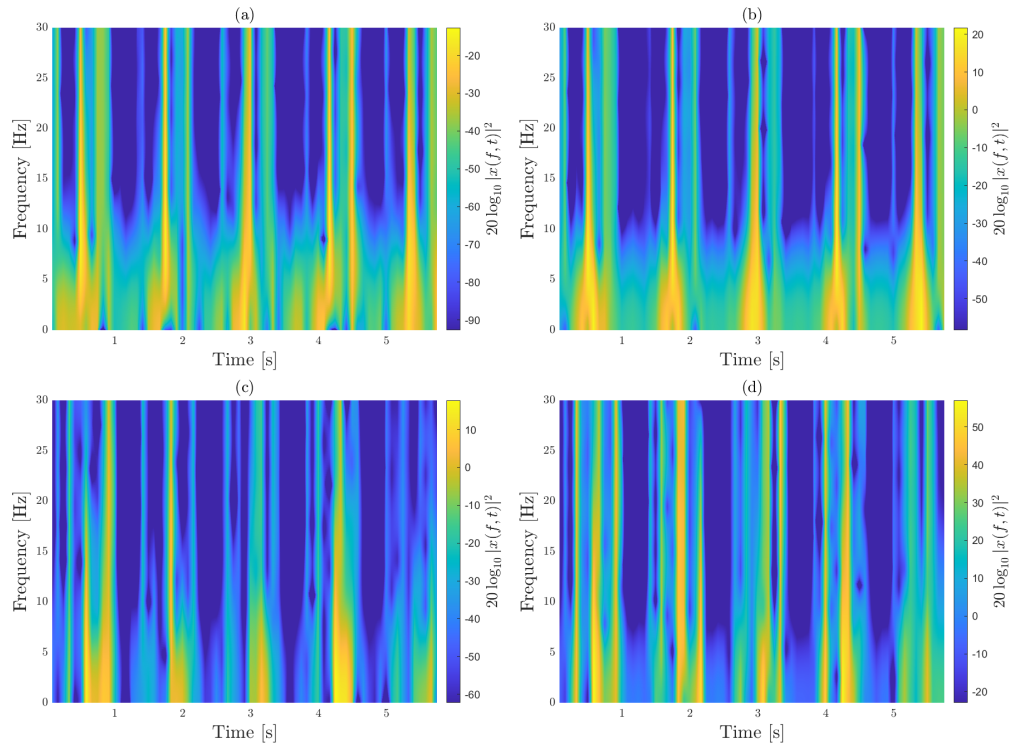


Fig. 4.37: Spectrogram of ICR trajectories during treadmill gait trials collected using the MVN Awinda system. (a) Shank ICR trajectory on the horizontal axis. (b) Shank ICR trajectory on the vertical axis. (c) Thigh ICR trajectory on the horizontal axis. (d) Thigh ICR trajectory on the vertical axis.

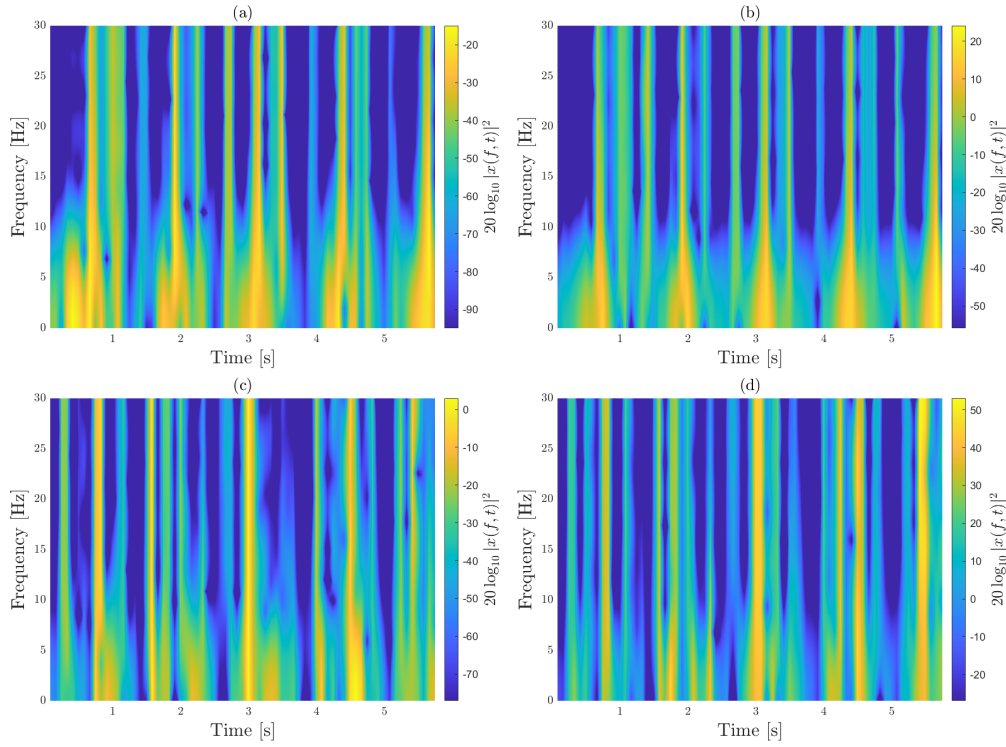


Fig. 4.38: Spectrogram of ICR trajectories during overground gait trials collected using the MVN Awinda system. (a) Shank ICR trajectory on the horizontal axis. (b) Shank ICR trajectory on the vertical axis. (c) Thigh ICR trajectory on the horizontal axis. (d) Thigh ICR trajectory on the vertical axis.

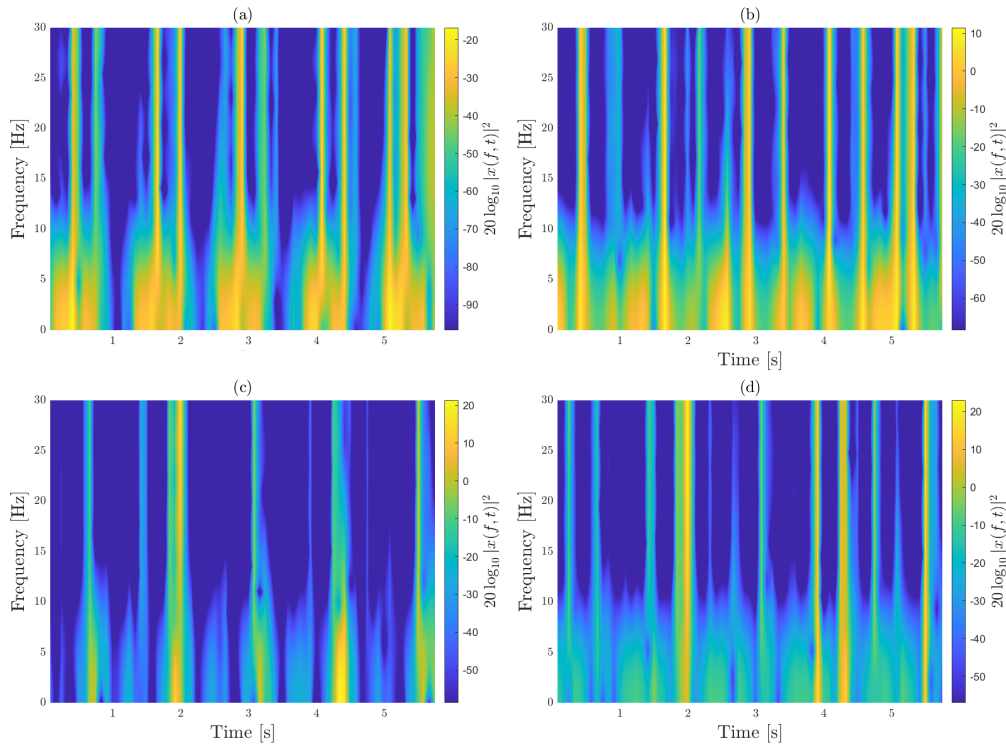


Fig. 4.39: Spectrogram of ICR trajectories during treadmill gait trials collected using Kinovea. (a) Shank ICR trajectory on the horizontal axis. (b) Shank ICR trajectory on the vertical axis. (c) Thigh ICR trajectory on the horizontal axis. (d) Thigh ICR trajectory on the vertical axis.

Figs. 4.34, 4.35, and 4.36 show how the PSD of the signal varies as a function of frequency when walking is considered a stationary process. In all cases, the dominant frequencies, identified as the PSD peaks, lie between 0 and 2 Hz. Furthermore, the plots indicate that many power peaks occur at frequencies below 10 Hz, which is consistent with the nature of the signal studied. ICR trajectories along the GC, with the exception of the moments of occurrence of the asymptotes, are characterized as a smooth signal with low oscillations.

Figs. 4.37, 4.38, and 4.39 show how the spectral power of the signal varies as a function of frequency and time (spectrograms). In this case, the gait is considered a nonstationary process. The spectrograms reveal asymptotes as vertical warm lines at specific time samples. This observation highlights a limitation of the STFT: even after truncating the time series magnitude to mitigate asymptotic effects, they still result in relatively high sustained power over all frequency bands at those time samples. This broad distribution of power complicates the precise identification of frequency components and challenges detailed spectral analysis. Moreover, although the magnitude is presented in decibels in order to visualize low-amplitude details in the spectrogram, apart from the fact that the signal power is contained mainly in low-frequency bins, not much more valuable information can be extracted.

By comparing the time series characteristics in the frequency domain, we can intuit that, contrary to expectations, there is a greater similarity between the data from the two gait trials collected with the MVN Awinda system than between, for example, treadmill gait trials collected with different data acquisition techniques. For this reason, special care must be taken when comparing data acquired with the MVN AWINDA system and vision-based systems during activities such as treadmill gait. When using the MVN Awinda system, the anatomical markers are always shifted forward as a result of the preprocessing performed by the Xsens engine, resulting in kinematic characteristics slightly different from the actual ones (obtained by vision-based methods). Due to this fact, it is recommended to use the system mainly to analyze the angular motion of the joints rather than the linear displacement of markers.

4.4.3 Analysis of ICR trajectories in state-space

According to the theory of systems, a system refers to an entity or object composed of interconnected components or parts that interact with each other (Backlund, 2000). This concept applies to all entities, whether they exist independently or as part of a larger structure. A system is called dynamic if it undergoes changes or evolves in behavior over time. Modeling dynamical systems involves the use of differential equations that describe the time evolution of the variables studied. Solving these differential equations is necessary to determine the precise response of the variable of interest to a given perturbation.

The formulation of the set of differential equations that model a dynamical system depends on the studied phenomenon; for instance, if what is analyzed is an electrical circuit, the equations are obtained applying the Kirchhoff laws, while in mechanical models, such as the musculoskeletal system, the applied theory corresponds to Newton laws and the strength of materials theory.

Dynamical systems can be significantly complex and characterized by multiple measurable variables, each with the potential to vary over time. To deal with multiple equations that span several orders, the state-space representation becomes a fundamental tool. This technique systematically describes the dynamics

of the system over time using a set of coupled first-order differential equations in a set of internal variables known as state variables, together with a set of algebraic equations that combine state variables into physical output variables (Hamilton, 1994). The concept of the state of a dynamic system refers to the minimum set of variables that fully describe the system and its response to any given set of inputs, so that knowing the current state of a given system and the effect on it of external perturbations allows predicting its future state. A set of differential equations expressed in matrix form using state-space representation follows the following expression.

$$\dot{x} = f(x, u, t) \quad (4.17)$$

where $x = [x_1, x_2, \dots, x_n]$ is the vector of state variables, $\dot{x} = [\dot{x}_1, \dot{x}_2, \dots, \dot{x}_n]$ is the vector of derivatives, and f is the rule that governs the evolution of the dynamical system.

Although the solution of the formulated differential equations allows us to accurately determine the state of the system at each instant of time, finding an exact solution can be complicated, especially when dealing with nonlinear systems. When modeling, linear differential equations are preferred due to the availability of exact solutions and the applicability of superposition rules. In cases where the equations are nonlinear, an attempt is made to approximate them to linear equations under assumptions made. An alternative approach to the study of the dynamics of systems, especially nonlinear dynamical systems, is to study their state-space, which is the space comprising the full range of possible states suitable for predicting the future evolution of the system (Bradley, 1999). This conceptual space serves as a global framework for understanding the dynamics of the system without the need to explicitly solve complex differential equations. Fig. 4.40 shows the phase portrait of a simple pendulum, which is the set of all possible trajectories of the simple pendulum in state-space, each given by a different initial condition.

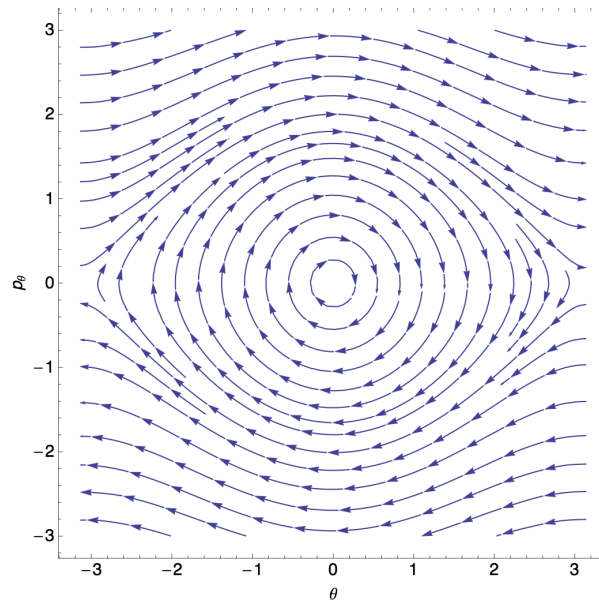


Fig. 4.40: Phase portrait of a simple pendulum.

The state-space of a dynamical system is usually studied by means of state-space plots, which facilitate the visualization of the trajectories traversed by the system as it evolves in time given initial conditions. This geometric tool is suitable for studying systems whose state is determined by up to three variables; however,

for systems of higher dimensions, it is necessary to examine their projections in a two- or three-dimensional space (Rodríguez-Licea et al., 2019). To construct a state-space plot, a coordinate is chosen for each state variable of the system, and the points are plotted at the corresponding set of coordinates of each observed state. Then, by joining the graphical points by straight lines (linear approximation), the trajectory followed by the system as it evolves in time starting from a given initial condition is obtained.

Time series acquired from experimental observations of a given dynamical system are usually insufficient to fully define the entire state-space and are considered projections of the system. However, according to Taken's theorem, measurements of a single state variable are sufficient to reconstruct the entire state-space of a dynamic system (Takens, 1981). To perform a reconstruction of the state-space, one of the commonly used methods is coordinate delay embedding, which consists of embedding the set of time series (measured by a sensor) into a higher-dimensional space, where each coordinate contains the time series with different delays (Vlachos and Kugiumtzis, 2008). Fig. 4.41 shows the principle of time delay embedding.

Let $[x_0, x_1, x_2, x_n]$ be the measured time series of a dynamical system, where x_i denotes the measured value at time i . The reconstructed state-space of the original system is given by the vector sequence:

$$y(i) = [x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{(m-1)\tau}] \quad (4.18)$$

where $y(i)$ refers to the reconstructed state-space at time i , m is the embedding dimension (dimension of the reconstructed state-space) and τ is the delay.

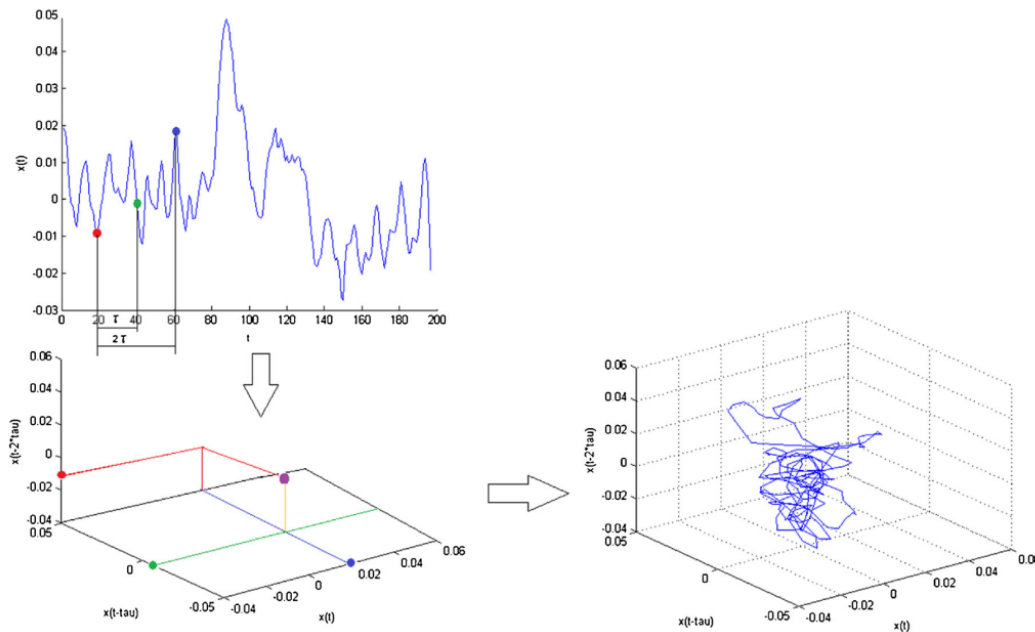


Fig. 4.41: Time delay embedding procedure. The coordinates of each point in the state-space are obtained as the values of the signal amplitudes at times $t = i$, $t = i + \tau$, and $t = i + 2\tau$.

When using coordinate delay embedding, two parameters must be chosen independently. The first parameter corresponds to the embedding dimension m , which determines the dimension of the reconstructed state-space, while the second parameter is the time delay. The Takens theorem establishes two sufficient (but not necessary) conditions to guarantee that the reconstructed phase state preserves the dynamics of the

observed system, given by (4.19) and (4.20). However, since the results obtained are highly dependent on the choice of these parameters, it is necessary to use sophisticated techniques to obtain adequate estimates.

$$\tau > 0 \quad (4.19)$$

$$m \geq 2d + 1 \quad (4.20)$$

where d is the dimension of the state space of the original dynamical system.

Among the techniques used for the estimation of the appropriate delay and embedding dimension, the most used techniques are the Average Mutual Information (AMI), and the False Nearest Neighborhood (FNN) (Marwan et al., 2007; Kantz and Schreiber, 2003). Within the AMI calculation, first the quantization of the time series is first carried out, in which its range is divided into many bins, which are used in the calculation of (4.21). The first minimum of $I(\tau)$ as a function of τ gives the optimal delay, since there we get the largest information by adding $x_i + \tau$.

$$I(\tau) = - \sum_{h=1}^{nBins} \sum_{k=1}^{nBins} P_{h,k}(\tau) \log \frac{P_{h,k}(\tau)}{P_h P_k} \quad (4.21)$$

where $nBins$ refers to the number of bins, P_h and P_k are the probability that an element of the time series x_i is in the h^{th} and k^{th} bin, respectively, while $P_{h,k}(\tau)$ is the probability that x_i is in the h^{th} bin while $x_i + \tau$ is in the k^{th} bin.

The use of FNN to estimate the embedding dimension, proposed by Kennel et al. (1992), is based on the following principle: Let us suppose that two one-dimensional time series data points are close together, so that they can be considered neighbors. If a single time series is embedded once (i.e., in two dimensions) using a certain delay τ , it is possible to use the coordinates of those data points to examine whether the distance between them has changed significantly. If the process of embedding the data in the reconstructed state space significantly changes the distance between neighboring points, these points are called false neighbors, and their presence indicates that it is necessary to embed the data in a higher dimension. Conversely, if their distance does not change appreciably, they are called true neighbors, and the current embedding dimension is sufficient, since the topology of the attractor has not changed significantly. Thus, the embedding dimension m can be estimated by examining the change in the distance between neighboring points in the state-space as we progressively embed the original time series in higher dimensions (Wallot and Mønster, 2018).

Let consider the embedding of a time series $x(t)$ in a state space of dimension m and denote by $y^{(r)}(t)$ the r^{th} nearest neighbor of the coordinate vector $y(t)$, then the distance between $y(t)$ and its r^{th} nearest neighbor is given by:

$$R_m(t, r) = \sqrt{\sum_{k=0}^{m-1} [x(t + k\tau) - x^{(r)}(t + k\tau)]^2} \quad (4.22)$$

where $R_D(t, r)$ is the euclidean distance calculated at time t between neighbors after embedding the time series $x(t)$ in a m dimensional space.

Applying the logic outlined above, embedding the time series $x(t)$ in a $(m + 1)$ -dimensional state-space adds a new coordinate to $y(t)$, hence the new distance between $y(t)$ and its r^{th} nearest neighbor is given by:

$$R_{m+1}(t, r) = \sqrt{R_m(t, r)^2 + [x(t + m\tau) - x^{(r)}(t + m\tau)]^2} \quad (4.23)$$

Within the musculoskeletal system, variables such as segment orientation, muscle length, and joint internal forces, among others, evolve over time. Rather than being fixed, these parameters exhibit changes as the system engages in various activities and movements. This dynamic nature allows the musculoskeletal system to respond and adapt to external forces, facilitating coordinated movements and functional capabilities. As a result, a convenient form of analyzing its behavior when walking is to see it as a dynamic system and employ techniques of the study of dynamical systems.

On the basis of the described process, the reconstruction of the state-space of the dynamical systems underlying the calculated time series was carried out. Because the time series in question are associated with the thigh and shank, two different approaches can be followed. The first is to consider each segment of the lower limb as an independent system and to analyze the state-spaces of each. The other approach is to consider all time series as coming from the same system and to analyze the state-space globally. Tables. 4.2 and 4.3 present the parameters chosen for the time delay embedding of the shank and thigh ICR trajectories when considered independent systems, and when considered as a whole, respectively. These parameters were obtained using the AMI and FNN.

Table 4.2: Estimated embedding parameters for the dynamic systems of the shank and thigh.

Lower limb segment	Gait trial	Embedding Dimension (m)	Time delay (τ)
Shank	Treadmill gait (MVN Awinda)	3	2
	Overground gait (MVN Awinda)	3	2
	Treadmill gait (Kinovea)	2	1
Thigh	Treadmill gait (MVN Awinda)	3	2
	Overground gait (MVN Awinda)	3	3
	Treadmill gait (Kinovea)	3	4

Table 4.3: Estimated embedding parameters for the dynamic shank-thigh system.

	Embedding Dimension	Time delay
Treadmill gait (MVN Awinda)	3	2
Overground gait (MVN Awinda)	3	2
Treadmill gait (Kinovea)	2	1

Once the embedding parameters are estimated, the time series is shifted in time by $k\tau$ samples for each new coordinate of the reconstructed m -dimensional space, where k is the number of the corresponding computed coordinate. However, as the reader can observe, Tables. 4.2 and 4.3, in all the cases considered, the embedding dimension is equal to or greater than 2, leading to the reconstruction of state spaces of dimensions equal to or greater than 4, since the data used for reconstruction comprise at least two time series (ICR trajectories on the x and y axes).

Having more than 4 dimensions is a problematic situation if one wants to track the trajectory of evolution of systems in state-space. When such cases occur, an alternative to analyze state-spaces is to perform projections of the state-space into two- or three-dimensional subspaces. However, the higher the dimension,

the more combinations of subspaces must be analyzed, making this alternative impractical. Another alternative is the use of a Recurrence Plot (RP), which allows investigating the dynamic trajectory of the system in high-dimensional state-spaces by means of a two-dimensional representation of its recurrences.

A recurrence in a dynamical system is its approximation, after a certain time, to a previous state. Recurrence is a fundamental property of deterministic dynamical systems and is typical of nonlinear or chaotic systems. Hence, one approach to analyze complex dynamical systems is to study their recurrences. RPs are a relatively new and powerful tool for advanced nonlinear analysis of dynamical systems. They are a two-dimensional representation of a square matrix R , where the elements of the matrix correspond to the times at which a state of a dynamical system recurs.

Recurrences of a state at time i at a different time j are marked within the matrix R with ones, while zeros are nonrecurring states. To facilitate visual interpretation of the matrix R , recurrences and nonrecurrences are represented by black and white dots on a 2D plot, where both axes are time axes. Therefore, the R matrix can be expressed as:

$$R(i, j) = \begin{cases} 1 : x_i \approx x_j \\ 0 : x_i \neq x_j \end{cases} \quad (4.24)$$

where x_i and y_i are states of the analyzed system.

The recurrences, $R(i, j) = 1$, are identified by defining a threshold distance ε , so that:

$$\|x_i - x_j\| \leq \varepsilon \rightarrow x_i \approx x_j \quad (4.25)$$

Observation of structures within RPs provides clues to the temporal evolution of the system. Those structures are presented on both large and small scales. Within the interpretation of the information contained in a RP, the large scale observations are refereed as typologies, while small-scale observations are known as textures.

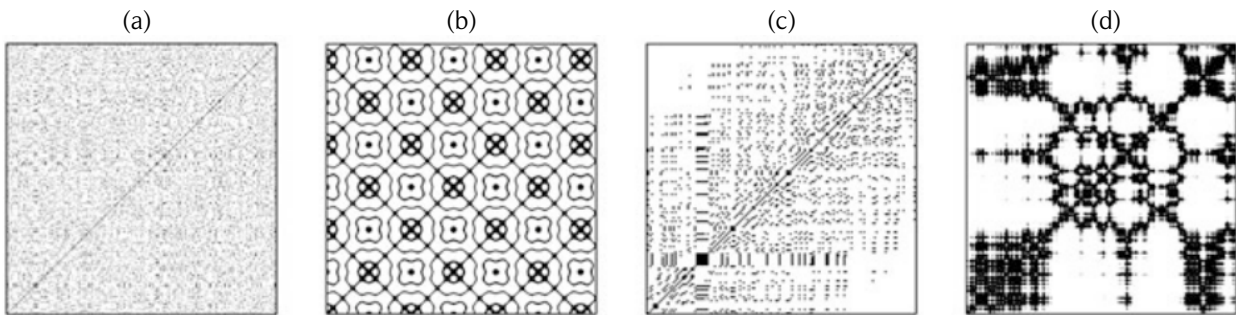


Fig. 4.42: Characteristic typology of RPs. (a) Homogeneous (uniformly distributed noise, $\varepsilon = 0.2$), (b) Periodic (superpositioned harmonic oscillations, $\varepsilon = 0.2$), (c) Drift ($\varepsilon = 0.2$), and (d) Disrupted (Brownian motion, $\varepsilon = 0.4$)

The large scale appearance of RPs, their typology, can be classified as homogeneous, periodic, drift, and disrupted:

- **Homogeneous RPs** typify stationary and autonomous systems in which relaxation times are short in comparison with the time spanned by the RP (see Fig. 4.42(a)).

- **Periodic RPs** have periodic or quasi-periodic structures such as diagonal lines or checkerboard structures and represent oscillating systems (see Fig. 4.42(b)).
- **Drifting RPs** are characterized by brightness in the upper left and lower right corners. They are obtained from systems with slowly varying parameters (nonstationary systems) (see Fig. 4.42(c)).
- **Disrupted RPs** are characterized by white areas or bands that represent abrupt changes in dynamics (see Fig. 4.42(d)).

At small scales (textures), it is possible to observe defined structures, such as isolated dots, dots forming horizontal, vertical, and diagonal lines, or dots grouped together to inscribe rectangular regions. Isolated recurrence points may occur if states are infrequent, if they do not persist for any length of time, or if they fluctuate greatly. Diagonal lines occur when one segment of the trajectory is parallel to another segment, i.e. the trajectory visits the same region of state-space at different times. The vertical and horizontal lines mark the length of time in which a state does not change or changes very slowly.

Table 4.4 presents a list of possible structures (typologies and textures) found in a RP and their corresponding qualitative interpretation:

Table 4.4: Summary of the interpretations of the structures found in the RPs.

Observation	Interpretation
Homogeneity	the process is stationary
Fading to the upper left and lower right corners	nonstationarity; the process contains a trend or drift
Disruptions (white bands)	nonstationarity; some states are rare or far from the normal; transitions may have occurred
Periodic/quasi-periodic patterns	cyclicities in the process; the time distance between periodic patterns (e.g. lines) corresponds to the period; long diagonal lines with different distances to each other reveal a quasi-periodic process
Single isolated points	heavy fluctuation in the process; if only single isolated points occur, the process may be an uncorrelated random or even anti-correlated process
Diagonal lines parallel to the main diagonal	the evolution of states is similar at different times; the process could be deterministic; if these diagonal lines occur beside single isolated points, the process could be chaotic
Diagonal lines orthogonal to the main diagonal	the evolution of states is similar at different times but with reverse time
Vertical and horizontal lines	some states do not change or change slowly for some time
Long bowed line structures	the evolution of states is similar in different epochs but with different velocity; the dynamics of the system could be changing

The RPs were obtained from the reconstructed state-spaces of each system considered. To define a recurrence, the Euclidean norm with $\varepsilon = 0.4$ was used.

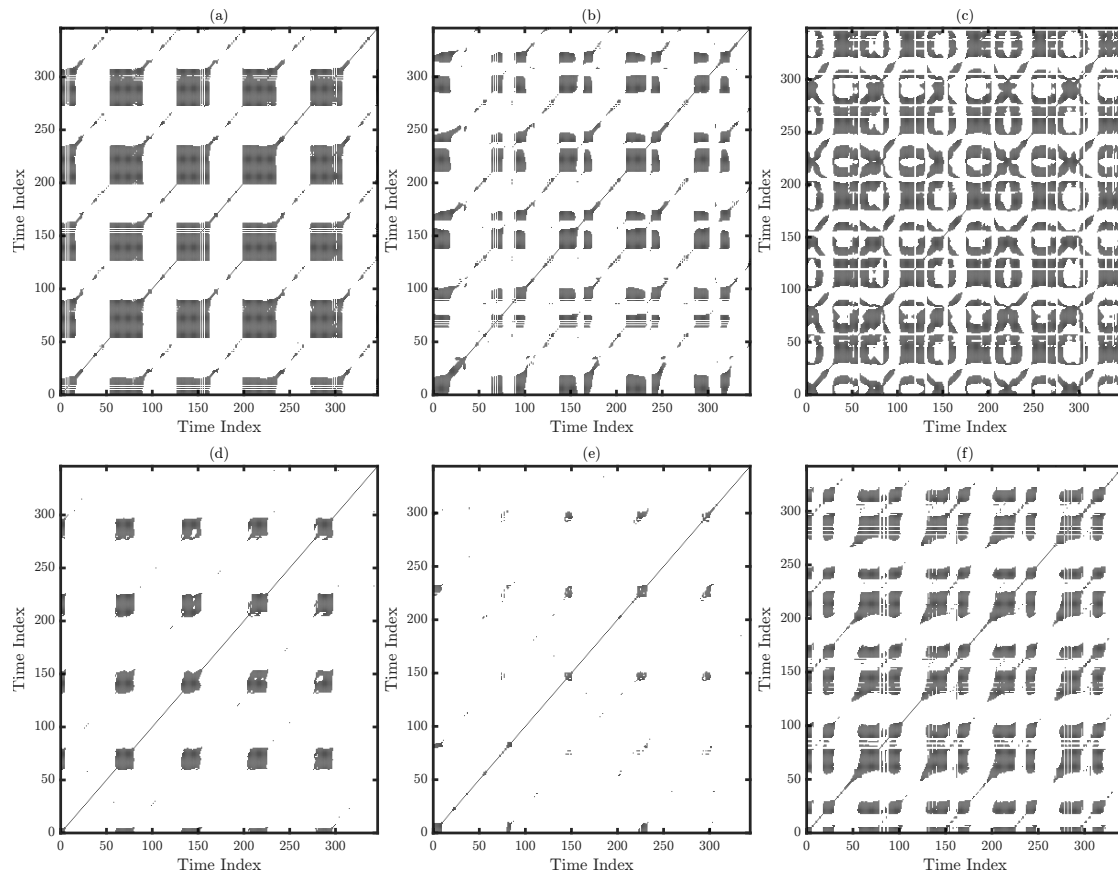


Fig. 4.43: RPs obtained from (a) shank ICR trajectory during treadmill gait trial using the MVN Awinda system (b) shank ICR trajectory during overground gait trial using the MVN Awinda system, (c) shank ICR trajectory during treadmill gait using Kinovea, (d) thigh ICR trajectory during treadmill gait trial using the MVN Awinda system, (e) thigh ICR trajectory during overground gait trial using the MVN Awinda system, and (f) thigh ICR trajectory during treadmill gait trial using Kinovea.

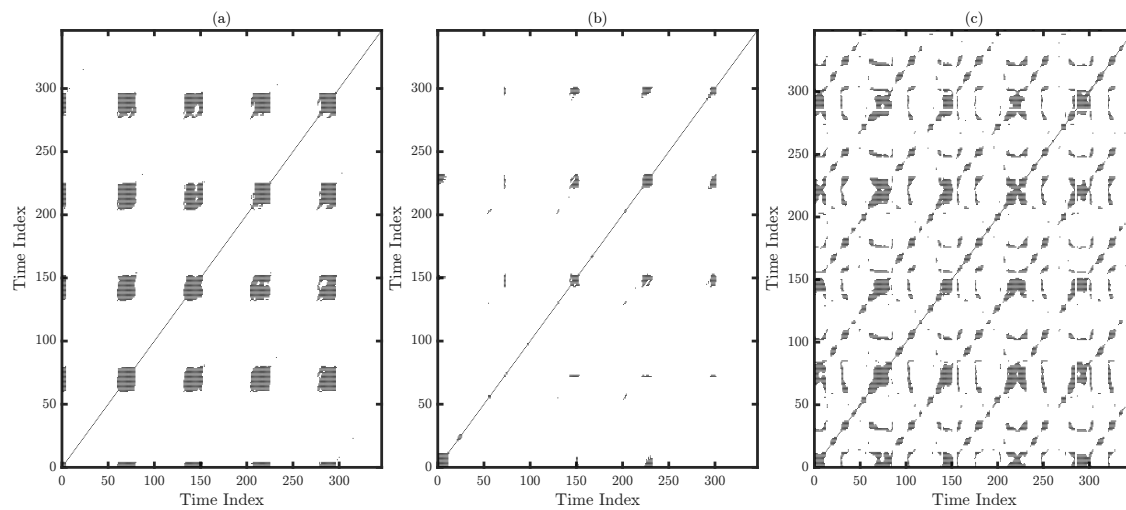


Fig. 4.44: RPs obtained from shank and thigh ICR trajectories during (a) treadmill gait trial using the MVN Awinda system (b) overground gait trial using the MVN Awinda system, (c) treadmill gait trial using Kinovea.

In each of the RPs obtained, repetitive structures are observed, typical of periodic systems or processes, although in each system these structures have a different shape. In the case of the RPs determined for the treadmill gait trial using the MVN Awinda system, both the shank and thigh systems contain structures similar to squares (see Figs. 4.43(a) and 4.43(d)). In the case of the RP determined for the shank system, the area of the squares is larger than in the RP determined for the thigh system, and diagonal lines are observed, typical of deterministic systems (see Fig.4.43(a)).

In the case of the overground gait trial, for the RP determined for the shank, squares cut by vertical and horizontal white lines are observed (see Fig. 4.43(b)). In addition, it is noticed that, although periodic, the shape of the structures presents slight variations; however, diagonal lines are always present. In the RP determined for the thigh system, small areas that are not clearly defined can be observed, but most of the plot corresponds to white regions, which implies that the system has few recurrences, suggesting a possible chaotic dynamic (see Fig. 4.43(e)). Continuing with the treadmill gait trial using Kinovea, in RPs we can observe a clearly periodic and well-defined complex structure for both the shank and the thigh system. The diagonal lines that characterize deterministic systems are present, mainly in the RP determined for the shank system (see Fig.4.43(c)), while in the RP determined for the thigh, although there are diagonal lines, they are less compared to the case of the shank system (see Fig.4.43(f)).

When analyzing the cases in which the shank and the thigh are considered a single comprehensive system, we can observe structures similar to those present in the RPs determined for the shank and thigh treated as independent systems. However, in all cases, the most present structures are those presented in the RP determined for the thigh system. This fact is especially noticeable in the RP determined for treadmill gait with the MVN Awinda system, in which small squares and the absence of diagonal lines are observed (see Fig.4.44(a)). In the RPs determined for the overground gait trials, we see that most of the plot is covered by white areas, as in the case of the thigh treated as an independent system (see Fig.4.43(b)). Finally, in the RP determined for the treadmill gait with Kinovea, we see a complex structure but less detailed than in the case of systems treated independently (see Fig.4.43(c)).

From the RPs determined for each system, we can conclude that the thigh system is the most feasible to use for gait analysis. The presence of diagonal lines in its structure indicates that it is a deterministic system that can be used to characterize gait. In the case of the thigh system, the graphs are full of white areas and isolated points showing that its states are infrequent, typical of a nonstationary, randomly correlated.

To provide an objective interpretation of the RPs, the Recurrence Quantification Analysis (RQA), developed by [Webber and Zbilut \(1994\)](#) and extended by [Marwan et al. \(2002\)](#), is often used. Basically, this analysis consists of numerically expressing the structures contained in the RP, which being possible through this method to identify temporal dependencies and transitions in the dynamics of the studied system. Some of the most commonly used RQA measures are presented below.

- Recurrence Rate (RR) (%): Percentage of recurrence points in an RP.

$$RR = \frac{1}{N^2} \sum_{(i,j)=1}^N R_{i,j} \quad (4.26)$$

- Determinism (DET) (%): Percentage of recurrence points that form diagonal lines.

$$DET = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=l_1}^N lP(l)} \quad (4.27)$$

- Longest Diagonal Line (L_{MAX}) (number of samples): Length of the longest diagonal line.

$$L_{MAX} = \max(l_i; i = 1, \dots, N_l) \quad (4.28)$$

- Shanon Entropy ($ENTR$) of the frequency distribution of the length of the diagonal lines (bits/bin):

$$ENTR = - \sum_{l=l_{min}}^N p(l) \log p(l) \quad (4.29)$$

- Laminarity (LAM) (%): Percentage of recurrence points that form vertical lines.

$$LAM = \frac{\sum_{l=v_{min}}^N vP(v)}{\sum_{l=v_1}^N vP(v)} \quad (4.30)$$

- Trapping Time (TT) (Number of samples): Average length of the vertical lines.

$$TT = \frac{\sum_{l=v_{min}}^N vP(l)}{\sum_{l=v_{min}}^N P(l)} \quad (4.31)$$

where l and v refer to the diagonal and vertical lines, N is the number of points in the state-space trajectory, N_l is the number of diagonal lines in the RP, and $P(l)$ and $P(v)$ are the histograms of the lengths of the vertical and horizontal lines, respectively.

In this study, the structures present in the RPs obtained for the thigh and shank systems and for the integral system formed by the thigh and shank were quantified using the described recurrence quantification measures. The results obtained are presented in Tables. 4.5 and y 4.6.

Table 4.5: RQA measures describing the structures in the RPs presented in Fig. 4.43

Gait Trial		RR	DET	L_{Max}	ENT	LAM	TT
Shank	Treadmill (MVN Awinda)	42.121	89.181	48.000	4.203	93.616	14.368
	Overground gait (MVN Awinda)	16.416	90.529	34.000	3.510	92.009	8.660
	Treadmill (Kinovea)	79.060	92.906	38.000	3.558	97.057	9.442
Thigh	Treadmill (MVN Awinda)	11.323	96.597	23.000	3.991	96.804	11.960
	Overground gait (MVN Awinda)	1.905	69.573	12.000	2.691	74.555	5.442
	Treadmill (Kinovea)	43.131	89.280	37.000	3.451	94.326	9.377

Table 4.6: RQA measures describing the structures in the RPs presented in Fig. 4.44

Gait Trial		RR	DET	L_{Max}	ENT	LAM	TT
Shank	Treadmill (MVN Awinda)	10.579	96.516	22.000	3.877	97.672	11.463
	Overground gait (MVN Awinda)	1.773	76.371	11.000	2.857	79.017	6.239
	Treadmill (Kinovea)	21.872	78.226	25.000	3.111	86.997	6.342

From the values obtained for each recurrence quantification measure of the systems considered, we can extract useful insights that are not evident in the RPs or that depend too much on the observer's perception. In Tables. 4.5 and 4.6, the first measure presented is RR . This measure quantifies the percentage of recurrent points that fall within the specified radius ε , and its value ranges from 0% (no recurrent points) to 100% (all recurrent points). Analyzing the values obtained from this measure, it is mainly confirmed that the thigh system recurs very little, which is indicated by its low values. This was also evident in the RPs characterized by a high percentage of white areas. The system with the lowest recurrences is the thigh, with a value of approximately 2% for the overground gait trials.

The values obtained for DET are high for all the systems considered, the highest value being 96.5% for the case of the thigh during the treadmill gait trial with the MVN Awinda system, from which we can conclude that all the systems have high probabilities of being deterministic. The lowest value of this measure was 69%, calculated for the thigh system in the case of the overground gait trial. From the values of L_{max} , we can state that the systems studied have a certain probability of being chaotic, especially the thigh system, whose value is 12 samples for the case of the overground gait trial. This statement is based on the fact that the shorter L_{max} , the more chaotic (less stable) the signal acquired from the system. The next calculated measure is the $ENTR$ of the frequency distribution of the length of the diagonal lines ($ENTR$), whose values for the systems considered are high, indicating the presence of high complexity and variation in system behavior. The maximum value of this measure was 4.2 bits/bin and was obtained for the shank system in the case of treadmill gait using the MVN Awinda system.

The last two measurements presented (LAM and TT) are associated with the presence of vertical lines in the RP, characterizing systems with states that are maintained over time or change slowly. In the systems studied, a strong presence of horizontal lines is observed due to the recurrent square structures observed in the RPs, with the most laminar structure observed in the shank system for the case of treadmill gait trial using the MVN Awinda system.

By comparing the Tables. 4.5 and 4.6, it can be concluded that combining the thigh and shank and analyzing the comprehensive system formed by the two segments implies in most cases dealing with a more chaotic and less deterministic system, which should be avoided. For this reason, subsequent analyzes will be limited to using only time series associated with the shank ICR trajectory in the sagittal plane as diagnostic gait data.

Chapter 5

A novel methodology to assess human gait by matching ICR trajectories of the segments of the lower limb in the sagittal plane

5.1 Introduction

This chapter, which constitutes the main part of the study, describes in detail the approach of a novel methodology for assessing human gait, as well as the theoretical assumptions on which its development is based. The proposed methodology uses measurements of the dissimilarity between the spatiotemporal kinematic characteristics associated with the movement of lower limb segments in the sagittal plane of different subjects to discriminate gait patterns. The assumptions for the formulation of the methodology were established on the basis of a detailed review of existing approaches to gait assessment, which is provided in Chapter 3. To refer to the proposed methodology, the codename ICR-LLS, which stands for Instantaneous Center of Rotation of the Lower Limb Segments, will be used. This methodology has the advantage of being strictly based on the analysis of gait-related kinematic data, which could be easily acquired under clinical conditions without the need for specialized equipment.

The ICR-LLS methodology relies on the study of the trajectories described by the ICR of the human shank in time and space, which differentiates it from traditional approaches found in the literature that are based mainly on the study of joint kinematics of the lower extremities. As a result of using the ICR-LLS methodology, a set of clusters is identified that partition the data set, grouping the elements (gait trials) that present similar gait patterns. Thus, subjects with similar gait patterns will belong to the same group, while subjects with dissimilar gait will belong to different groups. Using cluster analysis, we hope to be able to differentiate between subjects with gait disorders and control (healthy) subjects.

In addition to describing the proposed methodology, this section performs a sensitivity analysis to determine its suitability for discriminating patterns and its application in the diagnosis of gait disorders. For this purpose, additional treadmill and overground gait trials were performed using the MVN Awinda system,

during which different degrees of deviation from normal gait were simulated by adding a heel to the sole of the subjects' right foot only, thus adding difficulty to the gait and causing an asymmetrical gait pattern.

5.2 General principle

The ICR-LLS methodology, shown in Fig. 5.1, consists of four essential stages and one additional. In the first stage, the acquisition of gait-related kinematic data is performed using vision-based methods or inertial sensors. The acquired data correspond to the position and velocity in the sagittal plane of the centers of the knee and ankle joints. In the second stage, the position of the shank in the sagittal plane for the left and right lower extremities is determined using the collected data. This stage ends with the construction of a multivariate time series for each gait trial (performed by each subject). In the third stage, a pairwise comparison is performed between the multivariate time series using selected dissimilarity measures. This stage includes data pre-processing, in which time series are truncated, smoothed, and normalized. As a result of the third stage, the so-called dissimilarity matrix is obtained, whose inputs are pairwise comparisons between gait trials. Finally, in the fourth stage, a self-organization technique is used to represent the multivariate time series as points in a three-dimensional space, spaced according to the entries of the calculated dissimilarity matrix. As a result of this stage, a measure of spatial proximity or remoteness of the data is obtained from which it is possible to make inferences, such as differentiating groups that share similarities or recognizing outliers.

The last stage of ICR-LLS methodology is an optional stage that can be chosen to have an even more automated approach. In this stage, clustering analysis is used to identify within the data embedded in the 3D space those nodes that share similar characteristics (gait patterns). This stage is performed on the basis of the spatial density of the nodes within the space. Finally, the performance of clustering is assessed using different measures.

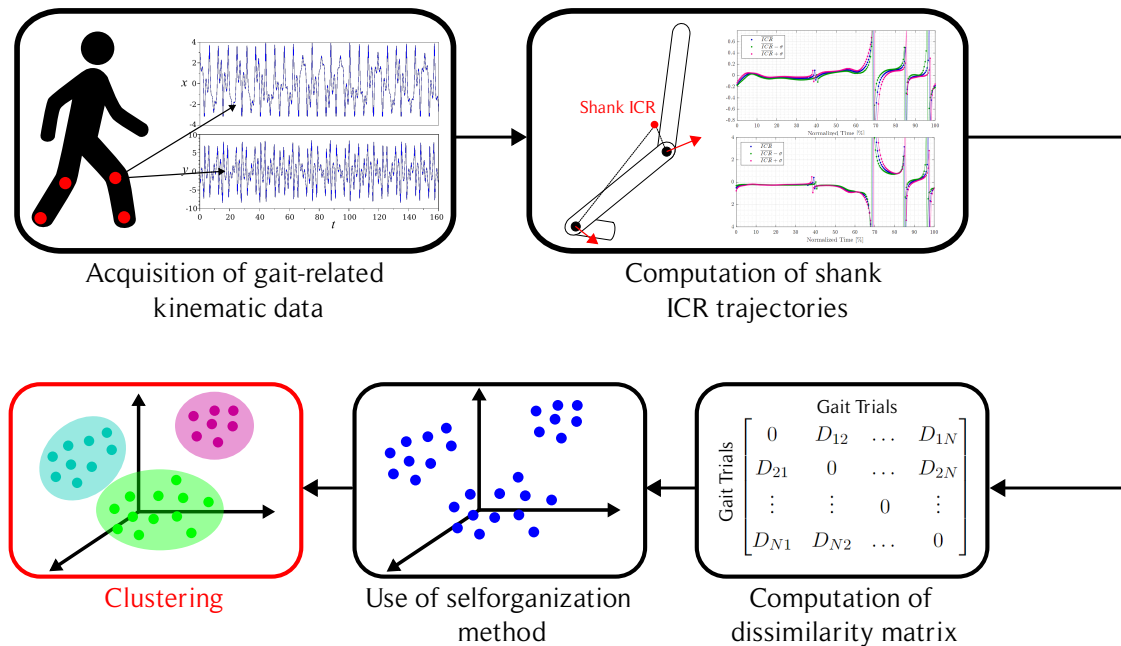


Fig. 5.1: Flow chart of the ICR-LLS methodology showing the sequential steps involved.

The proposed methodology can be used to assess various health conditions, such as neurodegenerative diseases and musculoskeletal injuries, as well as other factors affecting the subject's gait. This is based on the possibility that, among the identified clusters, those with a normal gait pattern can be grouped into the same cluster, and separated (spatially) from subjects with an abnormal gait pattern. In the best case, it would be possible to identify clusters composed only of subjects with abnormal gait.

5.3 Gait-related kinematic data acquisition and ICR computation

The ICR-LLS methodology requires as input data the position and velocity of the anatomical markers corresponding to the center of the knee and ankle joint on the horizontal and vertical axes of the sagittal plane of the subjects tested. Depending on the approach used to acquire the data, the velocity of the markers is obtained by numerical derivation of the marker position or directly provided by the MoCap.

The treadmill and overground gait trials were performed following the same data acquisition protocol presented in Chapter 4. The subjects who performed the gait trials were the same 15 subjects recruited for the previous study, whose purpose was to characterize the typical ICR trajectories of the tibia and thigh in the sagittal plane throughout the GC. For their demographics, the reader is referred to Table. 4.1. This new study considers only gait-related data collected with the MVN Awinda system. Although cameras were used to record motion in the sagittal plane of the lower extremities during treadmill gait trials, extracting kinematic data from these videos demands considerable processing time. The use of the MVN Awinda system is more feasible for handling large datasets, as it directly provides kinematic data, eliminating the need for additional user actions.

Each subject performed four gait trials on a treadmill and four on level ground. In both cases, in the first trial the subject walked freely, without restrictions or added difficulties. In the next three trials, a polyurethane heel was added near the subject's ankle to introduce different levels of gait difficulty (simulated gait abnormalities). The heel height was progressively increased by 1 cm in each trial, with the lowest level of difficulty being a 1.5 cm heel and the highest being a 3.5 cm heel. Therefore, the four treadmill and overground gait trials comprised normal gait (NG), level 1 heel gait (L_1), level 2 heel gait (L_2), and level 3 heel gait (L_3).

The data set used to test the proposed methodology included a total of 60 multivariate time series, each corresponding to the positions of the knee and ankle joint centers in Cartesian coordinates along the GC.



Fig. 5.2: Heel added to the sole of the subjects during gait trials to simulate different levels of gait abnormalities.

From the acquired data, the calculation of the ICR trajectory of the subject's shank along the GC is performed using the approach described in Chapter 4. First, the acquired data are segmented to isolate the individual strides performed by the subject. Then, the mean trajectory of the knee and ankle joint centers and their upper and lower limits during a stride are calculated. Finally, from the segmented trajectories, the position of the ICR of the shank is calculated, thus obtaining the mean ICR trajectory ($\overline{ICR}$), the upper limit ICR trajectory ($\overline{ICR} + \sigma$), and the lower limit ICR trajectory ($\overline{ICR} - \sigma$).

From each gait trial performed by a given subject, the resulting multivariate time series contains four time series that include the ICR trajectories of the left and right shanks on the vertical and horizontal axes. A single multivariate time series is structured in matrix form as follows:

$$S(i) = \begin{bmatrix} ICR_{rx}^{SH}(1) & ICR_{ry}^{SH}(1) & ICR_{lx}^{SH}(1) & ICR_{ly}^{SH}(1) \\ ICR_{rx}^{SH}(2) & ICR_{ry}^{SH}(2) & ICR_{lx}^{SH}(2) & ICR_{ly}^{SH}(2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ ICR_{rx}^{SH}(n) & ICR_{ry}^{SH}(n) & ICR_{lx}^{SH}(n) & ICR_{ly}^{SH}(n) \end{bmatrix} \quad (5.1)$$

where i refers to the corresponding gait trial of a given subject, ICR_{rx}^{SH} corresponds to the ICR trajectory of the left shank along the x axis, ICR_{ry}^{SH} corresponds to the ICR trajectory of the left shank along the y axis, ICR_{lx}^{SH} corresponds to the ICR trajectory of the right shank along the x axis, and ICR_{ly}^{SH} corresponds to the ICR trajectory of the right shank along the y axis.

Three multivariate time series are generated from each gait trial according to the structure present in (5.1). These time series contain $\overline{ICR}$, $\overline{ICR} + \sigma$, and $\overline{ICR} - \sigma$, respectively. This approach accounts for variations in the ICR trajectory of the shank across different gait cycles within a single gait trial.

5.4 Pairwise comparison of gait trials

In the ICR-LLS methodology, gait trials are represented by multivariate time series. A higher degree of variability observed between two time series is associated with a higher probability that one of the gait patterns compared presents some kind of anomaly. Therefore, pairwise comparisons are made between multivariate time series in the time domain. However, prior to pairwise comparisons, a pre-processing stage is performed in which the signals are cleaned, truncated, and normalized. After this stage, a selected measure of dissimilarity is used to quantify the dissimilarity between gait patterns.

5.4.1 Data preprocessing

Data preprocessing is essential to obtain high-quality information using data mining techniques. Its objective is to improve the relevance and appropriateness of the data, making it more suitable for analysis through the use of mathematical techniques and transformations. Each of the following described stages is applied to each computed shank ICR trajectory.

Handling of missing values

The first step in the data preprocessing stage is to ensure that the data are complete, which involves dealing with missing values in the data set that could significantly affect the accuracy of subsequent analyzes. In the case of time series, missing samples are not uncommon because of errors in the data acquisition process. In this study, data were acquired using the MVN Awinda system, which transmits the data acquired by the motion trackers to the computer at a fixed sampling rate of 60 Hz. To improve the quality of the data, the Xsens software performs an internal reprocessing step based on the built-in human biomechanical model (Schepers et al., 2018). This additional stage ensures high accuracy in the kinematic measurements and eliminates the problem of missing data.

Outliers detection

An important aspect to consider in the preprocessing stage is the presence of outliers in the data. Although outliers are usually the result of errors in data acquisition, they can occasionally represent unique features or anomalies that hold valuable information for diagnostic purposes. Therefore, in this study, it was decided not to eliminate outliers.

Data truncation

Recapitulating the results obtained in Chapter 4, the ICR trajectories of the shank exhibit asymptotes, which are samples that deviate significantly from their mean statistics. Comparing time series samples near an asymptote can lead to very misleading results if not handled carefully, as the differences between sample values can vary greatly, from 0 to infinity. Despite these variations, time series with similar patterns may still exist.

To avoid inaccurate time series comparisons, a data truncation step is applied. This step limits the range of values that the time series can reach. The upper and lower bounds are determined based on the typical mean characteristics observed in the shank ICR trajectories investigated in Chapter 4. Specifically, these bounds are placed symmetrically around 0, i.e., while the upper bound will have the value of a , the lower bound will have the value of $-a$. In computational terms, values outside the interval $[-a, a]$ are replaced by their nearest limit.

Data filtering

Although the Xsens software provides filtered data, it was decided to apply an additional filtration step using a Moving Average (MA) filter with a window length of 5 samples. The principle of the MA filter involves averaging a selected number of samples around a given data point and replacing the data point with the computed result (Schafer, 2011). This smoothing process helps remove high-frequency components from the time series, enhancing its overall quality. The operation of the MA filter is given by:

$$y(n) = \frac{1}{N} \sum_{m=0}^{N-1} x(n-m) \quad (5.2)$$

where $x(n)$ and $y(n)$ are the signal before and after filtering, respectively, and N is the length of the filtering window.

Data normalization

When comparing any type of data, a fundamental rule to keep in mind is that meaningful comparisons can only be made between objects of the same type. Similarly, as it is not feasible to compare apples with bananas, the same principle applies when comparing numerical data. To ensure meaningful comparisons between numerical data, an essential step is to bring the compared data series to the same scale. This transformation is known as normalization and consists of scaling the values to a predefined range, such as between 0 and 1. This process ensures that all data are presented within the same limits, eliminating the influence of different scales or units of measurement and facilitating direct comparisons (Fan et al., 2021).

A common method of normalization is standardization, which consists of subtracting the mean and dividing it by the standard deviation. This technique centers the data around zero and scales them according to their variability.

5.4.2 Pairwise comparison of time series using similarity measures

Similarity measures are quantitative measures used to assess the degree of similarity between two or more objects. These objects can take various forms of data representation, including scalar data, text, images, and time series. The quantitative relationship between a pair of objects is expressed through a single scalar number, often mapped to intervals such as $[-1, 1]$ or $[0, 1]$, where a value of 1 denotes the highest degree of similarity between objects (Cassisi et al., 2012). The following are similarity measures considered in this study:

Mean similarity

The mean similarity measure calculates the average similarity between the corresponding data points in two time series. It is often used to compare numerical data, where the similarity between two points is typically computed on the basis of the absolute difference. A higher mean similarity indicates a stronger resemblance between the data sets.

$$S_{Mean}(X, Y) = \frac{1}{N} \sum_{i=1}^N \left(1 - \frac{|x_i - y_i|}{|x_i| + |y_i|} \right) \quad (5.3)$$

where $S_{Mean}(X, Y)$ is the computed mean similarity, X and Y are two time series with the same number of data samples N .

Root mean square similarity

The Root Mean Square (RMS) similarity is a variation of the mean similarity metric that considers the squared differences between the corresponding data points. The RMS similarity is particularly useful when focusing on the overall differences between data sets. It measures the dispersion of the differences, giving

more weight to larger deviations.

$$S_{RMS}(X, Y) = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(1 - \frac{|x_i - y_i|}{|x_i| + |y_i|}\right)^2} \quad (5.4)$$

where $S_{RMS}(X, Y)$ is the computed RMS similarity, X and Y are two time series with the same number of data samples N .

Peak similarity

Peak similarity is commonly applied to time series data, where the aim is to identify similar peak patterns. This metric calculates the similarity between the corresponding peaks in two data sets, often using techniques such as peak detection and alignment.

$$S_{Peak}(X, Y) = \frac{1}{N} \sum_{i=1}^N \left(1 - \frac{|x_i - y_i|}{2 \max(|x_i|, |y_i|)}\right) \quad (5.5)$$

where $S_{Peak}(X, Y)$ is the computed peak similarity, X and Y are two time series with the same number of data samples N .

Cross-Correlation

Cross-correlation is a measure of the similarity of two time series that takes into account the time lag that may exist between them. It makes it possible to determine the degree of alignment of two time sequences, indicating the specific time at which the optimal alignment occurs.

$$S_{cross}(X, Y) = \max \left(\frac{\sum_{i=1}^N (x_i - \bar{X})(y_{i-k} - \bar{Y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (y_{i-k} - \bar{Y})^2}} \right) \quad (5.6)$$

where S_{cross} is the computed cross-correlation, $\bar{X}$ and $\bar{Y}$ denote the means of the compared time series X and Y , respectively, and k indicates the delay between them.

5.4.3 Pairwise comparison of time series using distance-based metrics

An alternative approach to comparing objects involves using distance-based metrics, which play a crucial role in addressing similarity issues in data mining tasks (Fu, 2011). Unlike similarity measures, distance-based metrics are not limited to a defined interval. A higher value of the metric indicates greater dissimilarity between the compared items, while a value close to zero indicates greater similarity.

Distance-based metrics must satisfy four fundamental properties: nonnegativity, identity, symmetry, and triangle inequality. If any of these properties are violated, the distance is considered nonmetric. Mathematically, these properties are expressed as:

$$f(x, y) \geq 0 \quad (5.7)$$

$$f(x, y) = 0 \Leftrightarrow x = y \quad (5.8)$$

$$f(x, y) = f(y, x) \quad (5.9)$$

$$f(x, z) \leq f(x, y) + f(y, z) \quad (5.10)$$

where x and y refer to the elements compared.

The use of a distance-based metric is favorable in many applications as it allows efficient indexing of the space to speed up search operations (Faloutsos et al., 1994). Distance-based metrics used in this study are presented below:

Euclidean distance

The Euclidean distance computes the straight-line distance between the corresponding data samples of two time series, providing the overall differences in magnitude and shape between the compared data. The Euclidean distance is particularly suitable for continuous and uniformly sampled data.

$$d_{Euclidean}(X, Y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (5.11)$$

where $d_{Euclidean}(X, Y)$ is the computed Euclidean distance and X and Y are two time series with the same amount of elements N .

Manhattan distance

Similar to the Euclidean distance, the Manhattan distance calculates the sum of the absolute differences between the corresponding data points. It is suitable for time series with high-dimensional and sparse data and is robust against outliers.

$$d_{Manhattan}(X, Y) = \sum_{i=1}^N |x_i - y_i| \quad (5.12)$$

where $d_{Manhattan}(X, Y)$ is the computed Manhattan distance, X and Y are two time series with the same quantity of elements N .

Dynamic time warping

Dynamic Time Warping (DTW) is a distance-based metric that allows to compare time series with varying time intervals or temporal distortions. Unlike traditional distance metrics, DTW nonlinearly aligns the time series, allowing for a more flexible and accurate comparison. The basic idea of DTW is to find the optimal alignment between the two time series that minimizes the distance between the corresponding data points. It is especially useful in situations where extensive prior knowledge is lacking or when the sequences have different lengths and varying time scales (Diab et al., 2019; Boulgouris et al., 2004).

Given two time series X and Y of length N and M , respectively, the alignment of X and Y using DTW exploits the information contained in the distance matrix d_{Matrix} , which has the following form:

$$d_{Matrix} = \begin{bmatrix} d(x_1, y_1) & d(x_1, y_2) & \dots & d(x_1, y_M) \\ d(x_2, y_1) & d(x_2, y_2) & \dots & d(x_2, y_M) \\ \vdots & \vdots & \ddots & \vdots \\ d(x_N, y_1) & d(x_N, y_2) & \dots & d(x_N, y_M) \end{bmatrix} \quad (5.13)$$

where $d(\cdot)$ denotes the distance used to compute the similarity between elements x and y of time series X and Y , respectively.

The optimal alignment between the time series is determined from the calculation of the optimal warping path, which is a sequence of indices representing pairs of aligned data points, indicating which points of one time series coincide with points of the other. The warping path contains contiguous elements of d_{Matrix} such that it minimizes the following function:

$$d_{DTW}(X, Y) = \min \left(\sqrt{\sum_{k=1}^K \omega_k} \right) \quad (5.14)$$

where $d_{DTW}(X, Y)$ is the computed DTW, and ω_k refer to the elements that form the warping path.

For the calculation of the warping path, dynamic programming is used (see Fig. 5.3). When applying this technique, the cumulative distance matrix is computed, which contains in each entry the following value:

$$\gamma(i, j) = d_{Matrix}(i, j) + \min [\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)] \quad (5.15)$$

where $\gamma(i, j)$ are the elements of the cumulative distance matrix.

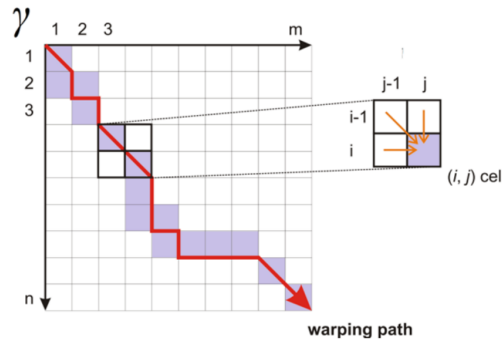


Fig. 5.3: Calculation of the warping path by using dynamic programming (Cassisi et al., 2012). Lavender cells correspond to the warping path, while the red arrow indicates its direction. The distance calculated in cell (i, j) takes into account, in addition to the distance $d(x_i, y_i)$, the minimum value between adjacent cells at positions $(i-1, j-1)$, $(i-1, j)$, and $(i, j-1)$.

Förstner distance

To quantify the extent to which a pair of coordinates changes together over time, covariance is often used. The covariance matrix is a square matrix that captures the covariance between each pair of variables in a multivariate data set. Its diagonal entries represent the variances of the individual variables, while the nondiagonal entries denote the covariances between the different variables.

Calculating the covariance matrix requires knowledge of the underlying probability distribution from which the samples $x(n)$ are drawn. However, this assumption is often unverifiable in real-world scenarios. As an alternative, the sample covariance matrix S is used. It serves as an unbiased estimator of the true covariance and offers the practical advantage that it can be computed using a finite number of samples (Cavazza et al., 2016). The sample covariance is defined by the following expression:

$$s(X) = \frac{1}{(N-1)} \sum_{i=1}^N (x_i - \bar{X})(x_i - \bar{X})^T \quad (5.16)$$

where X is the corresponding multivariate time series, N is its number of samples, and S is its computed covariance matrix.

The covariance matrix is widely used in data mining tasks. Its use has been documented in several applications related to human motion analysis, such as action recognition and gait evaluation (Li et al., 2018; Kumar and Babu, 2012). This mathematical element provides a feature representation that does not depend on time and therefore allows the extraction of a gait feature of constant size, regardless of variations in gait speed.

Based on the study proposed by Förstner and Moonen (2003), it is suggested that the similarity between covariance matrices should reflect the variance deviation in both directions according to the ratio of variances. Therefore, an appropriate distance-based metric for comparing covariance matrices is the sum of the squared logarithms of the eigenvalues, which is given by the below mathematical expression:

$$d_{Cov}(X, Y) = \sqrt{\sum_{i=1}^n \ln^2(\lambda_i(s(X), s(Y)))} \quad (5.17)$$

where $d_{Cov}(X, Y)$ is the computed Förstner distance of time series X and Y , $s(X)$ and $s(Y)$ are their respective covariance matrices, and $\lambda_i(s(X), s(Y))$ are the generalized eigenvalues that satisfy $s(X)x = \lambda_i s(Y)x$, with x being the corresponding generalized right eigenvector.

5.4.4 Ensemble of the dissimilarity matrix

Within the task of comparing time series, the similarity between two sequences of equal length can be assessed by summing the ordered point-to-point distances between the corresponding points. This approach is followed by most distance-based comparison techniques. However, it is not always possible to meet this requirement. Gait-related data may be acquired at different sampling rates, or recorded gait trials may vary in duration, and even when the data are acquired at the same sampling rate and the recordings last the same, errors in the acquisition process may result in differences in data length, making a point-to-point comparison infeasible. In such cases, alternative approaches, such as interpolation or alignment algorithms, may be necessary to allow meaningful comparisons between gait cycles (Lyons, 2010; Boulgouris et al., 2004).

In preparation for assembly of the dissimilarity matrix, the length of each time series is redefined using interpolation. The processed time series consists of 200 evenly spaced samples, regardless of the original length. Similarly, the timing within each time series was standardized so that the first sample corresponds to 0% and the last one to 100% of the GC.

The pairwise comparison measures used include both similarity and dissimilarity measures, meaning that while in some cases a value away from 0 corresponds to dissimilar items, in other cases a value away from 0 corresponds to similar items. Furthermore, the range of values obtained from the use of similarity measures is scaled in an interval from 0 to 1, while in the case of distance-based metrics, there is no upper limit. With this in mind, it was decided to transform similarity measures into dissimilarity measures using (5.18). The values obtained are scaled between 0 and 50, so all the metrics used fall between the same limits.

$$d(X, Y) = \frac{1}{s(X, Y)} \quad (5.18)$$

where $d(X, Y)$ is the calculated dissimilarity measure from the similarity measure $S(X, Y)$.

Comparison of multivariate time series was performed using each described measure, and the structure of the assembled dissimilarity matrices is as follows::

$$M = \begin{bmatrix} 0 & D_{12} & \dots & D_{1N} \\ D_{21} & 0 & \dots & D_{2N} \\ \vdots & \vdots & 0 & \vdots \\ D_{N1} & D_{N2} & \dots & 0 \end{bmatrix} \quad (5.19)$$

where each matrix entry $D_{i,j}$ is the pairwise comparison between gait trials i and j .

The diagonal of the dissimilarity matrix contains zero values, since it corresponds to the comparison between a gait trial and itself. On the other hand, the matrix is symmetric with respect to its diagonal, i.e., each entry $D_{i,j}$ is equal to $D_{j,i}$.

Given that three multivariate time series are obtained from each gait trial, which are $\overline{ICR}$, $\overline{ICR} + \sigma$, and $\overline{ICR} - \sigma$, the assembly of the dissimilarity matrix involves two main steps:

1. Computing pairwise dissimilarity measures between multivariate time series of the same type, for instance between $\overline{ICR}$ s.
2. Averaging the three resulting dissimilarity matrices.

From this process, a global dissimilarity matrix is obtained that accounts for the variation in the shank ICR trajectories between gait cycles within the same gait trial.

5.5 Self-organization method for data embedding in 3D space

The computed dissimilarity matrix provides information on the dissimilarity between the gait patterns of subjects at different difficulty levels. A simple way to interpret this matrix is to imagine that each gait trial corresponds to a node located at a specific location in the plane or space and that the element of the matrix refers to the distance between pairs of nodes. However, even if we could hypothetically embed the gait trials in a 2D or 3D space, it would still be geometrically impossible to obtain precise coordinates for each node that would completely agree with the information contained in the dissimilarity matrix. That is, the dissimilarity matrix contains a certain degree of inconsistency.

This incoherence increases with the number of elements in the matrix and decreases with the higher dimension of the space in which the nodes are embedded. Ideally, one would like to embed the data in the highest dimension space possible to mitigate this problem. However, it is impractical to embed the data in a space of dimension greater than 3, since it is not possible to observe the arrangement of the nodes graphically.

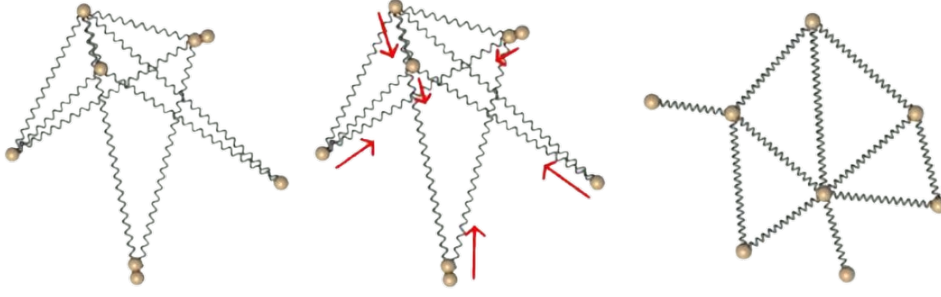


Fig. 5.4: Force-directed graph layout drawing technique (Kobourov, 2012).

The next step of the ICR-LLS methodology consists precisely in embedding the gait trials in a three-dimensional space from the information contained in the dissimilarity matrix. In a three-dimensional space, the set of gait-related data constitutes a network, where each node corresponds to a different gait trial, and the length of the edges is proportional to the dissimilarity between gait trials. To find the geometric layout of the network, a self-organizing method based on the force-directed graph layout drawing technique is used. This technique, also known as the electric spring model, calculates the layout of a network using only the information contained in the network structure itself, rather than relying on domain-specific knowledge (Yang and Liu, 2013). The principle of operation of the force-directed graph layout drawing technique is based on the assignment of attractive and repulsive forces between the nodes of the network. Two nodes connected by an edge experience both attractive and repulsive forces, while two unconnected nodes experience only repulsive forces. These forces, determined by the distances between the nodes, drive the nodes into a network configuration with a minimum total energy (Hu, 2006). An analogy to understand the method is to imagine that we are dealing with a mechanical system, where the nodes are steel rings and the edges are the springs, as shown in Fig. 5.5.

In this study, all nodes, corresponding to gait trials, are connected. Consequently, all pairs of nodes receive attractive and repulsive forces. This configuration ensures that nodes with higher dissimilarity repel each other, while those with lower dissimilarity (or higher similarity) attract each other.

Let $G(V, E, K)$ be an undirected weighted network, where V is the set of nodes, E is the set of edges, and W is the set of weights (i.e., the measure of dissimilarity) on each edge. The assigned repulsive force and the attractive force are given by (5.20) and (5.21), respectively.

$$f_r(i, j) = \frac{-CK^2}{\|y(i) - y(j)\|} e^{\alpha|D(i,j)|} \quad (5.20)$$

$$f_a(i, j) = \frac{\|y(i) - y(j)\|^2}{K} e^{\frac{\gamma}{|D(i,j)|}} \quad (5.21)$$

where C regulates the strength of the repulsive force relative to the attractive force, K is the 'natural' distance between nodes for which nodes encounter no attractive or repulsive forces (called 'natural spring length'), and α and γ are tuning parameters.

The combined force resulting from the interaction between node i and j is defined by the combination of (5.20) and (5.21) which results in the following expression:

$$f_c(i, j) = f_r(i, j) + f_a(i, j) \quad (5.22)$$

where $f_c(i, j)$ is the combined force acting between nodes i and j .

The function to be minimized is the total energy of the system, which is calculated as the sum squared of the forces acting on each of the nodes. This process is carried out through several iterations, in which the nodes move towards a position of lower energy (Meyers et al., 2020). As the network energy converges, the nodes self-organize in 3D space, forming a unique topology and producing new node positions that preserve the edge weights between nodes, i.e., the dissimilarity distances between gait patterns (Kobourov, 2012).

It is important to clarify that the generated network layout does not necessarily assimilate all the information contained in the dissimilarity matrix. Instead, it preserves the information implicit in the dissimilarity matrix in an acceptable proportion, allowing the geometric distances between nodes to be computed, which are more coherent distances than those present in the dissimilarity matrix. After self-organization, clusters of nodes that share similar patterns emerge. The relative distances between nodes in the network reflect the dissimilarities between subjects' gait patterns, providing valuable insights from these clusters. Consequently, node locations can be used as input variables in clustering or classification models to predict possible gait disorders.

Using a self-organizing method brings three important advantages to the proposed methodology: (1) it reduces the dimensionality of the data by transforming the time series into three-dimensional time-independent nodes, (2) it allows to visually represent the distribution of the time series in a subspace of reduced dimensions, and (3) it reduces the inconsistency in the dissimilarity matrix.

5.6 Clustering of gait-related data

The next step of the ICR-LLS methodology, which is an optional step, involves identifying within the generated network the clusters of nodes with similar patterns using clustering analysis. Cluster analysis, a key task in data mining, divides the data into a series of batches or specific groups, so that the data belonging to the same group have similar properties and the data belonging to different groups have dissimilar properties (Ergüner Özkoç, 2021). It comprises a wide range of methods based on different principles, including K-Means, Affinity Propagation, Mean-shift, DBSCAN, Gaussian Mixtures, and Spectral Clustering, among others, which excel in several applications (DeCann et al., 2014; Zgolli et al., 2018; Alqahtani et al., 2021).

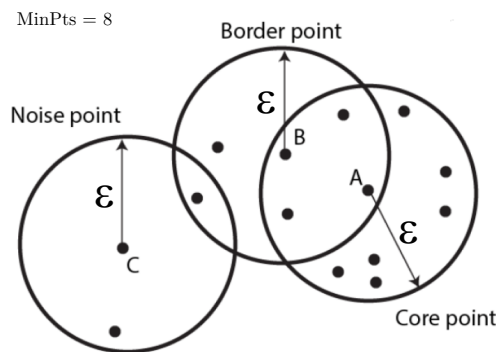


Fig. 5.5: DBSCAN point types [Adapted from Kanwal and Asghar (2021)]

In this study, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering algorithm was used. The idea behind its operation is that clusters are dense regions in the data space separated

by low-density regions (Daszykowski et al., 1996). To determine whether a region has a high data density, DBSCAN verifies for each element of the data space whether its vicinity (sphere of defined radius) contains at least a minimum number of points. Its main advantage over more basic algorithms, such as K-means, is that it can detect clusters of arbitrary shapes in the presence of noise and does not need the number of clusters as prior knowledge (Gao et al., 2023).

The DBSCAN algorithm requires two parameters. The parameter ε defines the size of the neighborhood around a data point. That is, if the distance between two points is less than or equal to ε , then they are considered neighbors. If the value of ε is chosen too small, a large part of the data will be considered low-density areas; on the contrary, if it is chosen too large, the clusters will merge and most of the data points will be in the same clusters. The second parameter (*MinPts*) is the minimum number of neighbors within the neighborhood of a point.

The DBSCAN algorithm consists of the following steps:

1. A value is chosen for ε and *MinPts*.
2. The distances between all pairs of data points are calculated.
3. Data points are classified as core points, border points, and noise points according to the following rules. The core points are points that have more than *MinPts* points in their neighborhood. Border points are points that are not core points but in whose neighborhood there is a core point. The rest of the points are noise points (see Fig. 5.5).
4. Noise points are eliminated.
5. All core points located at a distance ε from each other are assigned to the same cluster.
6. Border points are assigned to clusters associated with the core points in the neighborhood in which they are located.

The idea of using clustering analysis in the ICR-LLS methodology is to assist in the recognition of similar patterns in the generated network by automatically assigning in the same group gait trials that belong to the same category, e.g. healthy control subjects. However, this step may not be necessary if the network explicitly shows a high density of nodes in easily distinguishable regions.

Clustering performance is evaluated using two different metrics. For the internal evaluation of the clusters, the Silhouette Coefficient (SC) is used, while for the external evaluation, the Normalized Mutual Information (NMI) is used. In the internal evaluation of clusters, it is assumed that the actual labels of the data are unknown, i.e., it is not known with certainty which categories are present in the data to be clustered. The silhouette metric measures the cohesion of nodes within each cluster, providing information about their proximity, while assessing the separation between clusters (Řezanková, 2018). The silhouette coefficient $s(i)$ of an individual sample i is given by:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (5.23)$$

where a is the average distance between a sample and all other points in the same cluster, and b is the average distance between a sample and all other points in the nearest cluster.

The values of the silhouette coefficients near 1 indicate that the nodes exhibit high cohesion and separation. Values close to 0 indicate that the nodes are more or less evenly distributed around the cluster centers. Finally, values close to -1 indicate that nodes may have been assigned to erroneous clusters.

In the external evaluation, it is necessary to know the actual labels of the data, i.e. the clusters to be formed are known, and it is desired to check whether the algorithm used is able to identify them correctly. NMI is based on the concept of entropy, which quantifies the amount of uncertainty in the distribution of cluster assignments in a given cluster. The entropy of a given cluster U is given by:

$$H(U) = \sum_{c=1}^k -p(U_c) \log(p(U_c)) \quad (5.24)$$

where $p(U_c)$ is the proportion of nodes belonging to the cluster U_c

Mutual information between clusters U_1 and U_2 is given by:

$$MI(U_1, U_2) = \sum_{c=1}^k \sum_{l=1}^m p(c, l) \log \left(\frac{p(c, l)}{p(c)p(l)} \right) \quad (5.25)$$

where $p(c, l)$ is the proportion of nodes belonging to both clusters U_1 and U_2 .

NMI between clusters U_1 and U_2 is given by:

$$NMI(U_1, U_2) = \frac{2MI(U_1, U_2)}{H(U_1) + H(U_2)} \quad (5.26)$$

Since NMI values fall on a scale of 0 to 1, comparisons can be made between clustering with different numbers of clusters (Amelio and Pizzuti, 2015). NMI values closer to 1 indicate that the formed clusters are closer to the ground truth (actual clusters).

5.7 Results and discussion

The results of applying the ICR-LLS methodology to distinguish the different grades of added gait difficulties are presented below. The first part corresponds to a sensitivity analysis, where comparisons of the gait of the same individual subjected to different levels of gait difficulty are made to assess the methodology's potential to identify deviations from the normal pattern. The second part analyzes whether, given a data set containing several subjects with different levels of gait difficulty, the methodology is able to recognize clusters of subjects with the same level of gait difficulty.

5.7.1 Sensitivity analysis

The results of the sensitivity analysis performed are presented below.

From the acquired data, the set of multivariate time series corresponding to the ICR trajectory of the shank in the sagittal plane was calculated for each of the 15 subjects under the four different conditions of gait difficulties (NG , L_1 , L_2 , and L_3) and for the two types of gait trials (overground and treadmill).

Figs. 5.6 and 5.7 show the left shank ICR trajectory throughout the GC obtained from treadmill gait trials and overground gait trials, respectively.

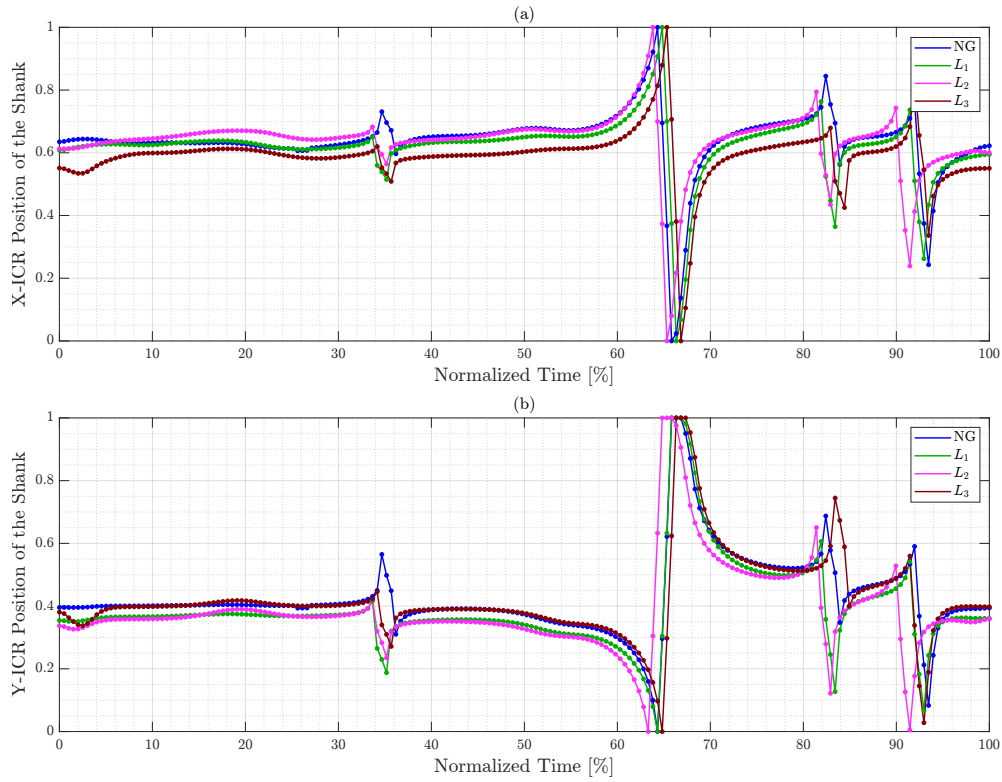


Fig. 5.6: Shank ICR trajectories during treadmill gait trials for different levels of gait difficulty. (a) Horizontal axis. (b) Vertical axis.

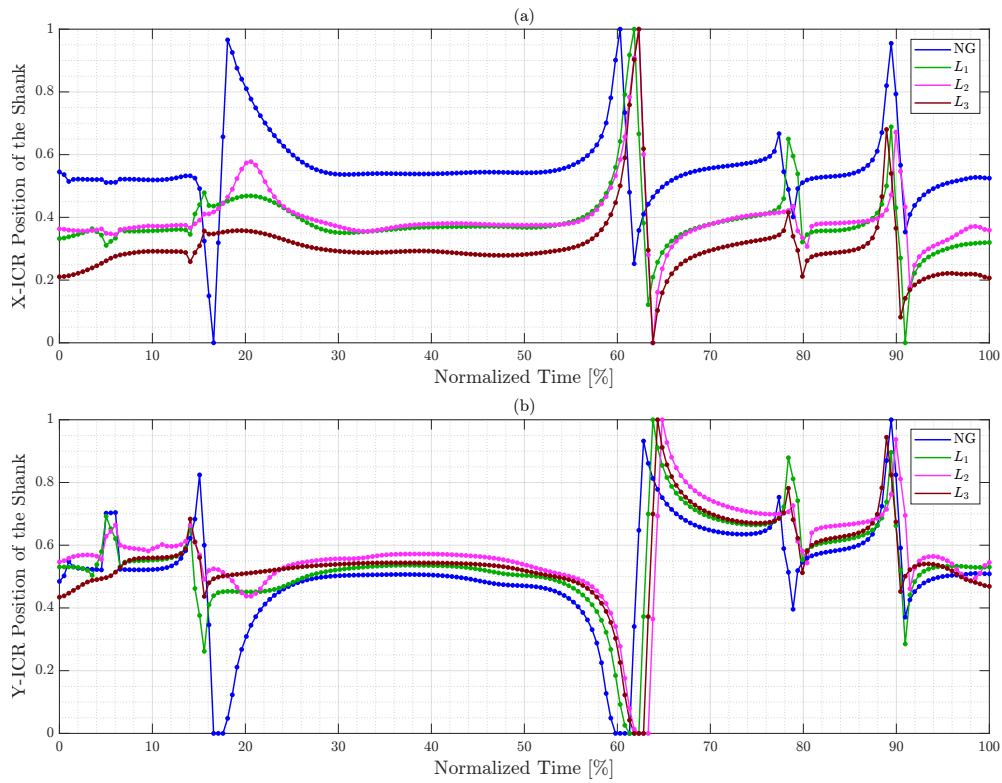


Fig. 5.7: Shank ICR trajectories during overground gait trials for different levels of gait difficulty. (a) Horizontal axis. (b) Vertical axis.

As can be seen, the distinct (averaged) ICR trajectory of the shank on the x and y axes, discussed in Chapter 4, is still evident despite the transformations applied during preprocessing. However, preprocessing can result in sequences of consecutive samples in the time series with the same value. This phenomenon is illustrated in Fig. 5.6(b), where several samples around 65% of the GC possess the maximum value. This occurs when a rapidly ascending asymptote is present, causing nearby samples to reach values that exceed the bounds. The truncation method replaces these outbound values with the corresponding limit value, leading to the observed effect.

Shank ICR trajectories calculated from data acquired from NG gait trials and those calculated from data acquired from gait trials with added difficulty share the same characteristic pattern. Specifically, the shank ICR trajectories obtained from treadmill gait trials appear to resemble the same time series but shifted in magnitude. Larger deviations are observed between the different levels of gait difficulty for the ICR trajectories obtained from treadmill gait trials. Moreover, there appears to be a higher similarity between the gait trials L_1 , L_2 , and L_3 than between these and NG . This suggests that use shank ICR trajectories to distinguish patterns is feasible, at least between normal gait and simulated cases of abnormal gait.

After preprocessing, the multivariate time series are compared using dissimilarity measures. The measures used include the eight described in Section 5.4, i.e., Euclidean distance, Mean similarity, Peak similarity, Cross-correlation, DTW, Förstner distance, Manhattan distance, Root mean square, and a new metric that consists of initially applying cross-correlation to align the time series with the best match and then calculating DTW.

Figs. 5.8 and 5.9 present the dissimilarity matrix resulting from the pairwise comparison between the four gait type trials on treadmill and overground, respectively. The calculated dissimilarity matrices correspond to an average of the dissimilarity matrices obtained for the different subjects who performed the gait trials with different levels of added difficulty. Averaging the dissimilarity matrices seems to be a logical step if the data extracted from the gait trials performed by each subject are considered as occurrences of the same process. To facilitate the interpretation of the dissimilarity matrices, it was decided to scale each of the dissimilarity metrics so that the lower boundary value was 0 and the upper boundary value was 50. In addition, the matrices are presented using color maps, where the colors tend to become warmer as the calculated value of the dissimilarity metric increases. Thus, cells that tend more towards red represent very dissimilar pairwise comparisons, while blue cells represent very similar pairwise comparisons.

A remarkable fact of the structure of the dissimilarity matrix is that its main diagonal contains values equal to zero, represented by dark blue cells, since the cells that make up it contain the comparison between a time series and itself.

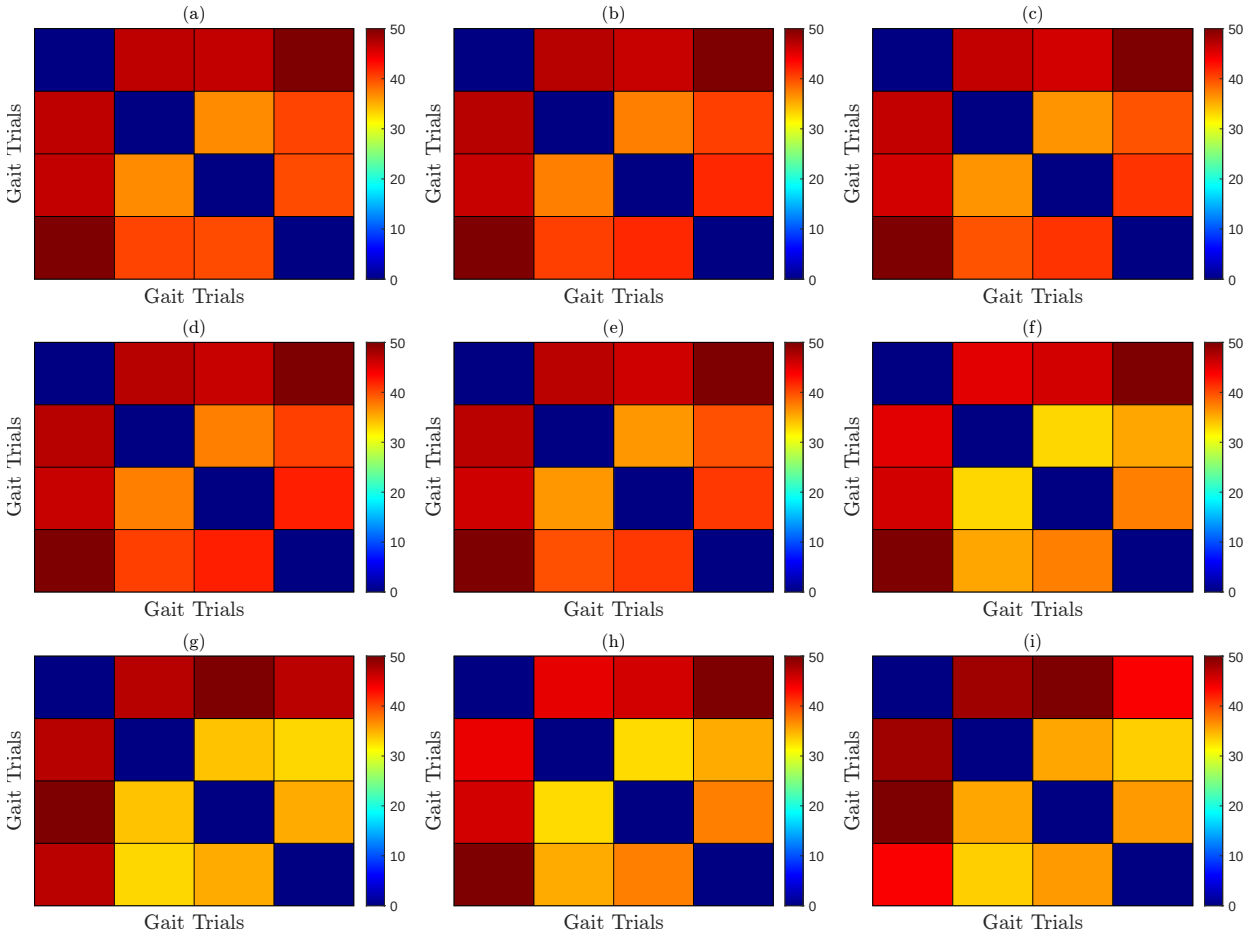


Fig. 5.8: Dissimilarity matrices calculated from pairwise comparisons of shank ICR trajectories of the same subject during treadmill gait trials using (a) Euclidean distance, (b) mean dissimilarity, (c) RMS dissimilarity, (d) peak dissimilarity, (e) Manhattan distance, (f) DTW, (g) inverse cross-correlation, (h) cross-correlation + DTW, and (i) Förstner distance. In each matrix, ordered from top to bottom and from left to right, there are gait trials corresponding to NG , L_1 , L_2 and L_3 , respectively.

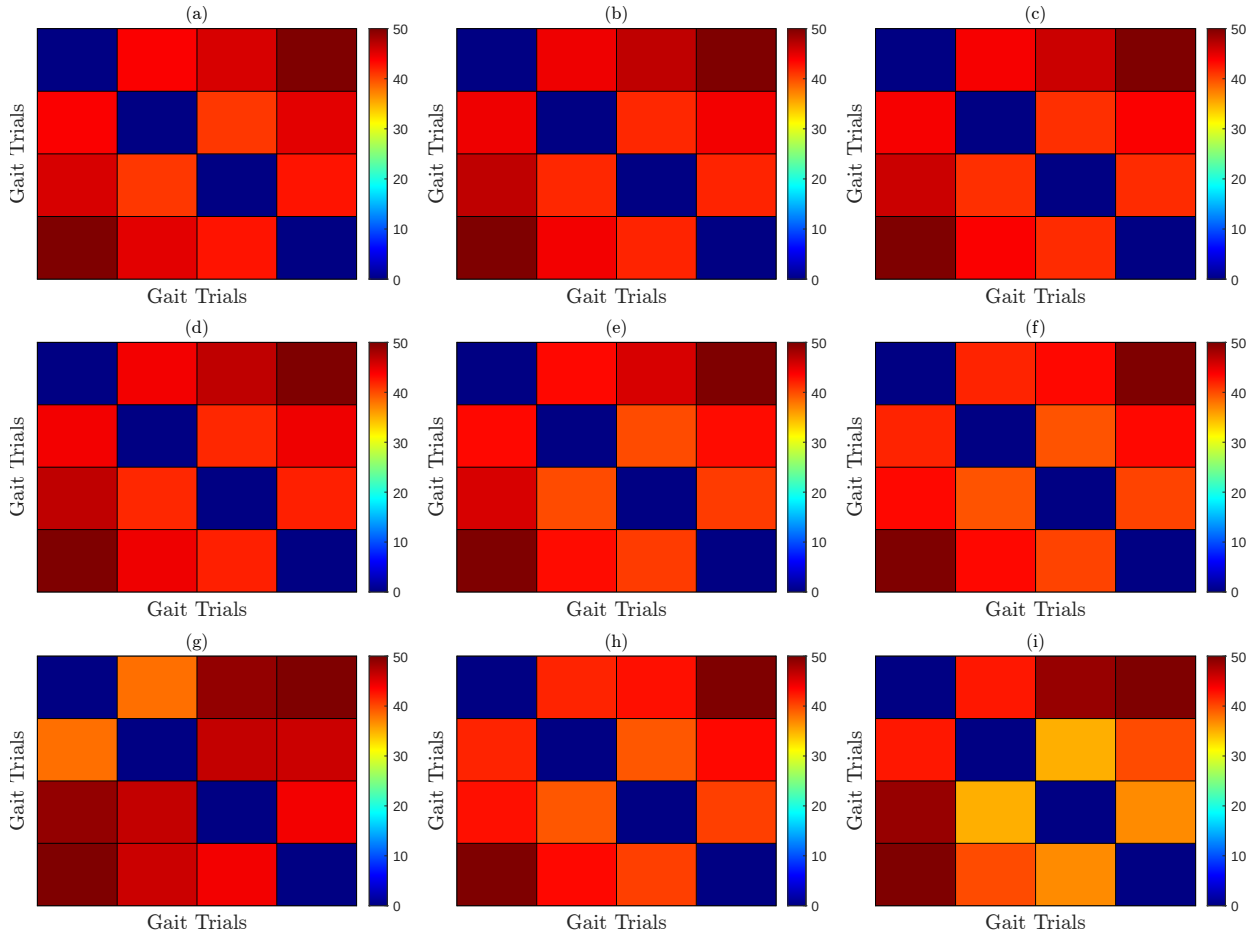


Fig. 5.9: Dissimilarity matrices calculated from pairwise comparisons of shank ICR trajectories of the same subject during overground gait trials using (a) Euclidean distance, (b) mean dissimilarity, (c) RMS dissimilarity, (d) peak dissimilarity, (e) Manhattan distance, (f) DTW, (g) inverse cross-correlation, (h) cross-correlation + DTW, and (i) Förstner distance. In each matrix, ordered from top to bottom and from left to right, there are the gait trials corresponding to NG , L_1 , L_2 and L_3 , respectively.

In the dissimilarity matrices concerning treadmill gait trials, we can observe that the highest values are found in the first column and first row, which is a positive result considering the purpose of the proposed methodology, as this means that the normal gait trial is further away from the gait trials with added difficulty. Regarding the dissimilarity matrix concerning the overground gait trials, most of its cells contain values very close to the maximum, so the differences between the normal gait and the gait with added difficulty are less evident. However, in the dissimilarity matrix calculated from the Förstner distance, in which the greatest color contrast is observed (see Fig. 5.9(i)), the values of the cells appear to increase as we move from left to right and from top to bottom, which is positive because it indicates that the dissimilarity between normal and simulated abnormal gait is greater as the subject walks with a higher degree of added difficulty.

From the dissimilarity matrices, the self-organization method based on the electric spring model is used to generate a network containing the four gait trials performed by each subject. To generate the network, a simulation consisting of 8000 iterations was performed, enough for the algorithm to converge and for the relative positions between nodes to change minimally if the iterations are continued.

Figs. 5.10 and 5.11 show the resulting network arrangement obtained from each of the dissimilarity matrices regarding treadmill and treadmill trials, respectively.

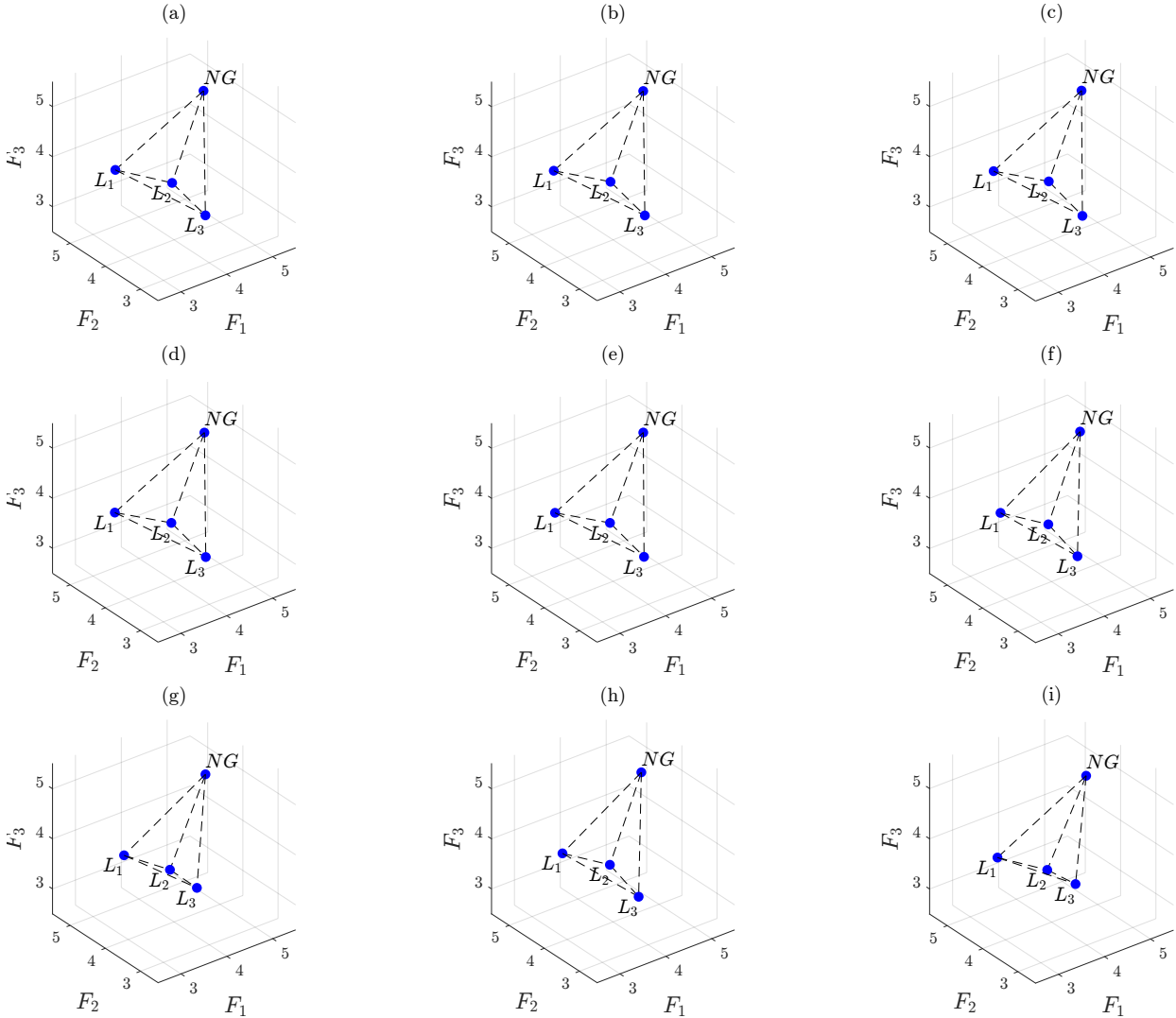


Fig. 5.10: Networks generated from data collected during treadmill gait trials of the same subject at different difficulty levels. The networks layout was calculated using the information contained in the dissimilarity matrices calculated based on (a) mean similarity, (b) RMS similarity, (c) Euclidean distance, (d) Manhattan distance, (e) cross-correlation, (f) peak similarity, (g) DTW, (h) Förstner metric, and (i) cross-correlation + DTW

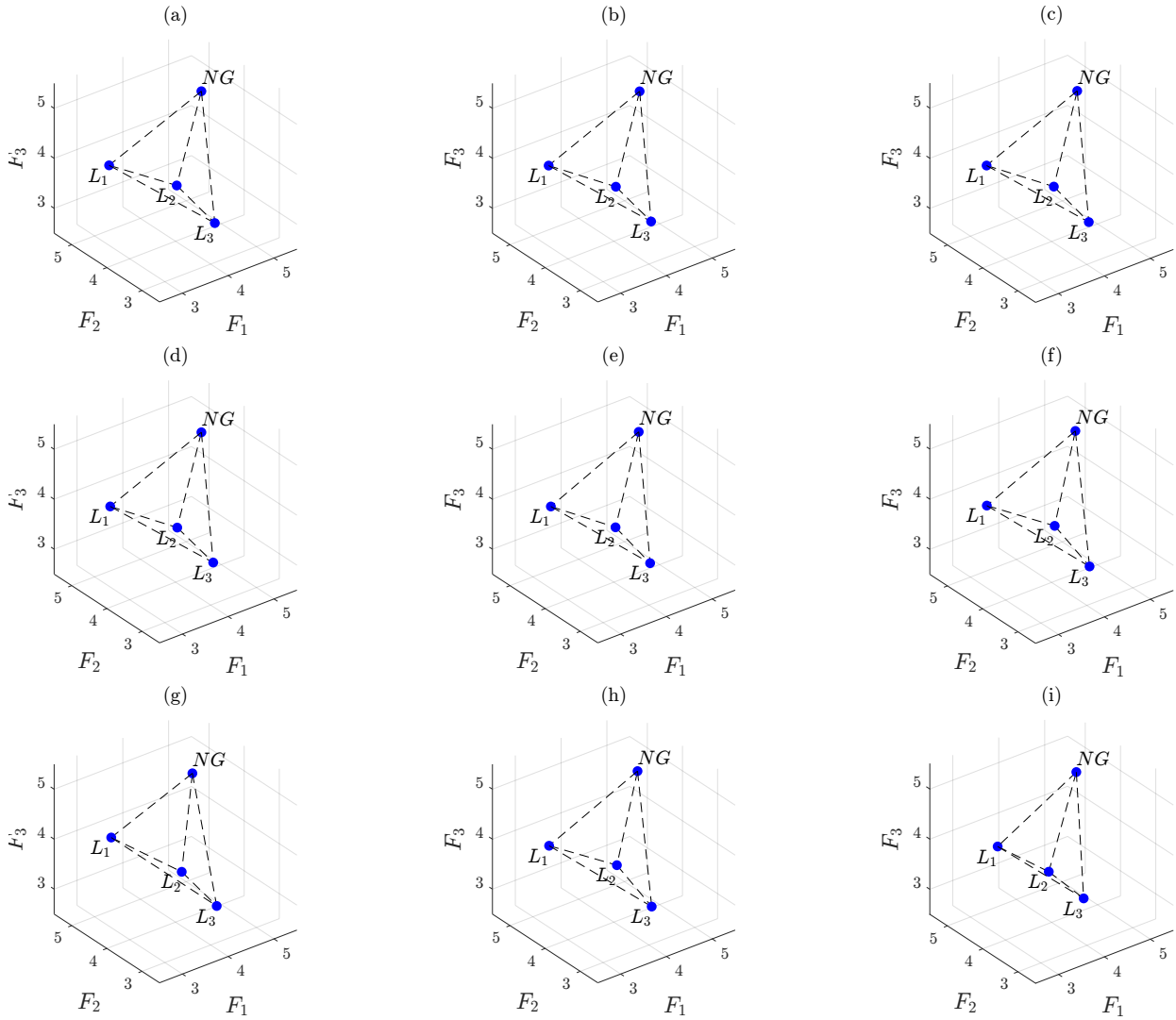


Fig. 5.11: Networks generated from data collected during overground gait trials of the same subject at different difficulty levels. The network layout was calculated using the information contained in the dissimilarity matrices calculated based on (a) mean similarity, (b) RMS similarity, (c) Euclidean distance, (d) Manhattan distance, (e) cross-correlation, (f) peak similarity, (g) DTW, (h) Förstner metric, and (i) cross-correlation + DTW

It can be seen that the layout of the different networks is very similar. The nodes are arranged in a pyramid whose longest edges correspond to the edges between NG and the other nodes, with the exception of the network. This indicates that the normal gait NG has the highest degree of dissimilarity within pairwise comparisons between the gait trials, which had previously been directly inferred from the dissimilarity matrices. However, the network shown in Fig. 5.11(g) seems to be an exception, as the node corresponding to the L_3 gait appears to be the farthest away.

Since the networks generated by the force-directed self-organization method are three-dimensional graphs, each node of the network is defined by three coordinates. That is, each gait trial was transformed from a multivariate time series to a vector defined by only three elements that characterize the gait.

Figs. 5.12 and 5.13 show the value of the coordinates of the nodes, which will be referred to as gait features, for all generated networks. Within the operation of the self-organization method, no preferred coordinates are chosen on which the nodes tend to move more. Each coordinate is independent and has the same level of relevance. Moreover, the graph could be rotated, the axes could be interchanged, and the information contained in the graph would not change. The location of one node relative to another is the useful information extracted from the network and not its coordinates with respect to the origin.

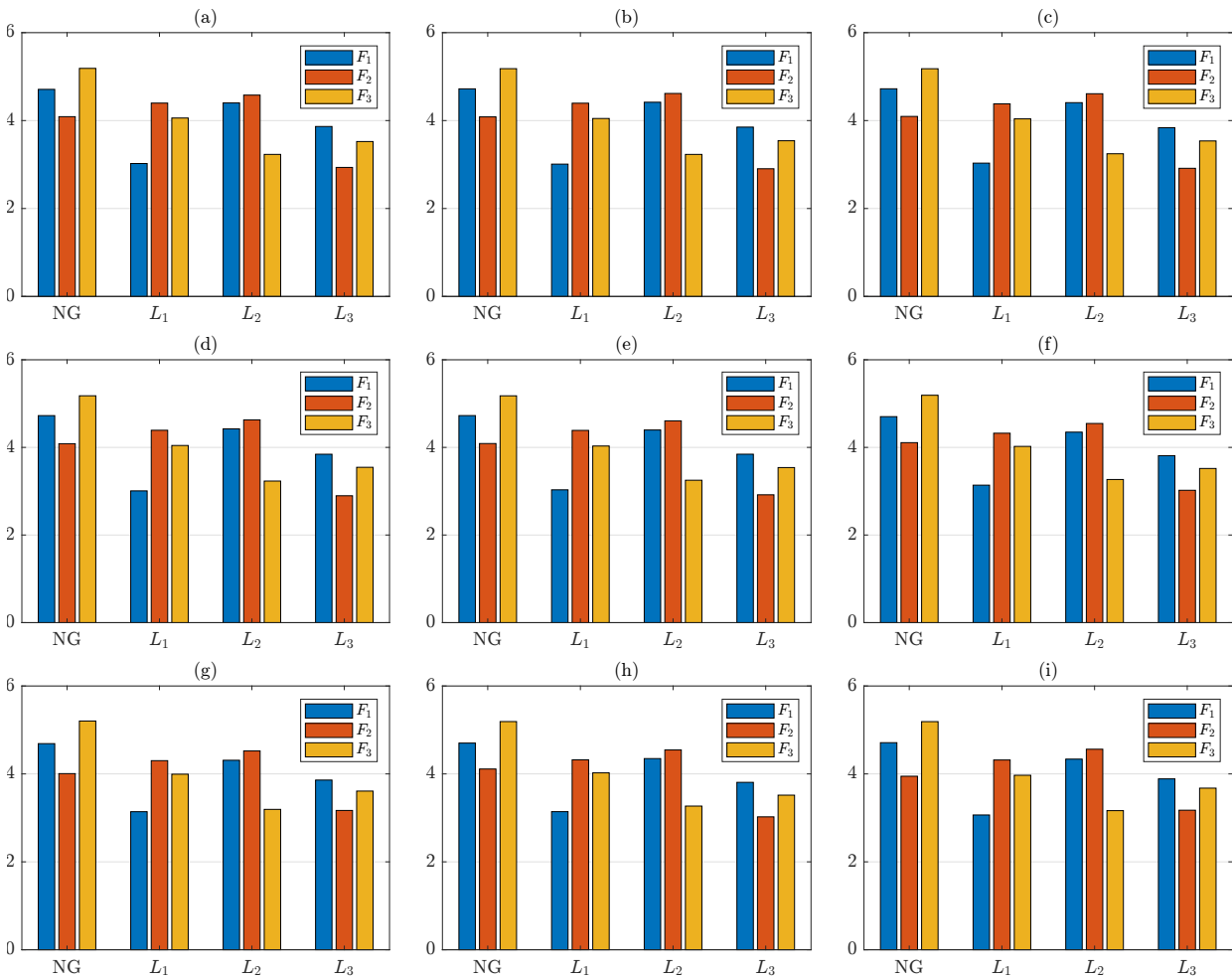


Fig. 5.12: Extracted gait features from data collected during treadmill gait trials at different difficulty levels. Gait features were obtained from the dissimilarity matrices computed by using (a) mean similarity, (b) RMS similarity, (c) Euclidean distance, (d) Manhattan distance, (e) cross-correlation, (f) peak similarity, (g) DTW, (h) Förstner metric, and (i) cross-correlation + DTW

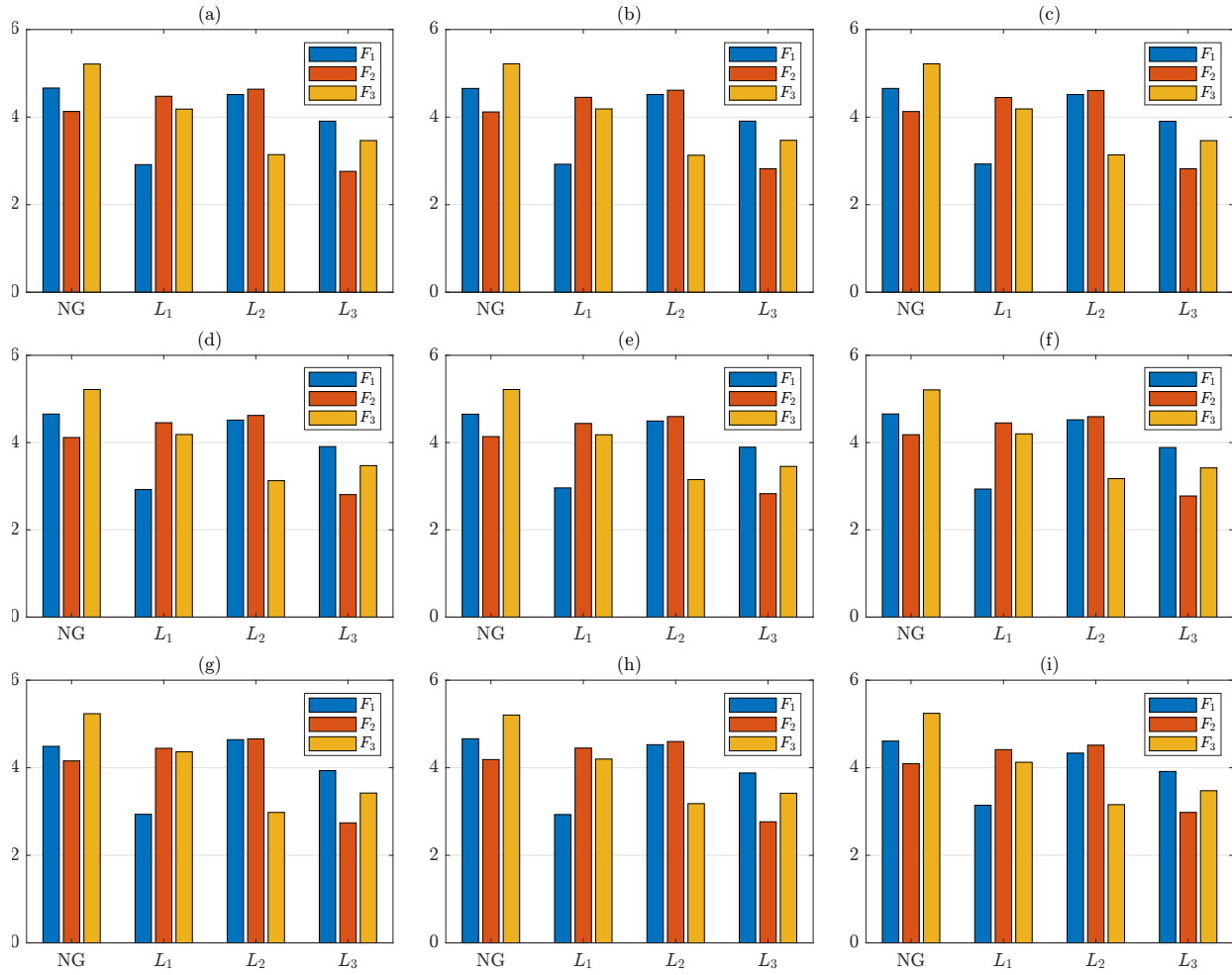


Fig. 5.13: Extracted gait features from data collected during overground gait trials at different difficulty levels. Gait features were obtained from the dissimilarity matrices computed using (a) mean similarity, (b) RMS similarity, (c) Euclidean distance, (d) Manhattan distance, (e) cross-correlation, (f) peak similarity, (g) DTW, (h) Förstner metric, and (i) cross-correlation + DTW

In Figs. 5.12 and 5.13, we can observe that the bar corresponding to the gait feature F_3 , which in the network is the z -axis, takes the highest value in the gait trial NG , and its value decreases as the difficulty of the gait increases. However, its value increases again in the gait trial in which the highest level of difficulty is added (L_3), which is not expected. The other two gait features do not appear to have a definite tendency to increase or decrease. Another notable result is that both the F_1 and F_3 gait features are maximal in most of the generated networks. This pattern seems to be the same for both the treadmill and ground gait trials, at least in the case of the subject studied.

The study of the correlation between the gait features extracted from the networks and the degree of gait difficulty is the approach chosen to analyze the sensitivity of the proposed methodology. However, since there are three factors that can vary (three gait features) and each correlates differently and independently, no clear conclusion is obtained as to how the outcome of the methodology varies with higher levels of gait difficulty.

To obtain a more comprehensive measure, a simple two-step procedure is proposed. In the first step, the gait features extracted from each network (computed using different dissimilarity measures) are averaged.

Consequently, the gait features are examined in a single bar chart that comprises the effect of all the similarity measures used to calculate the dissimilarity matrices. Fig. 5.14 shows the results of performing this step.

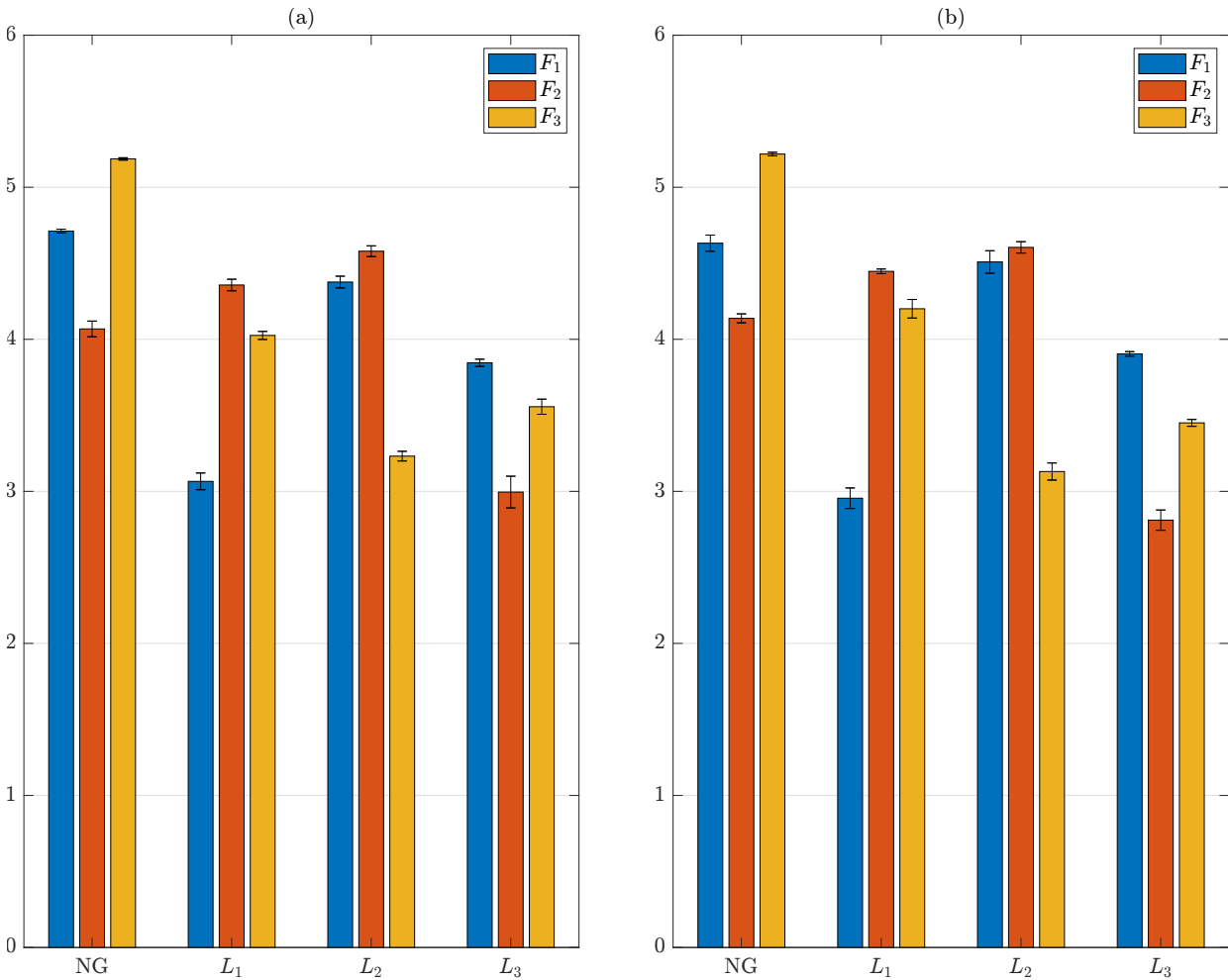


Fig. 5.14: Average gait features extracted from data collected during (a) treadmill gait trials and (b) over-ground gait trials. For their calculation, the gait features extracted from the different networks layout generated were averaged.

Since the layout of the networks, generated based on the different dissimilarity matrices, is in general similar, so are the values of the gait features (node coordinates) and, therefore, the obtained bar charts are similar to the independent bar charts presented in 5.12 and 5.13 for the treadmill and overground gait trials, respectively. Comparing Figs. 5.14(a) and 5.14(b), we can notice that the gait features follow the same trends in both cases (treadmill and overground) and that there are only small variations in the length of the bars. The bar charts also provide information on the standard deviation of gait features, which in all cases is considerably small, not exceeding a value of 0.12.

The three gait features that characterize a gait type are orthogonal (uncorrelated) variables; therefore, in the next step a characteristic value is calculated for each gait type using the Euclidean distance. In other words, the norm calculated from the three gait features is used to represent a gait type, which is advantageous, since a single scalar defines a gait type. Furthermore, before conducting this step, the value of the gait

features characterizing normal gait (NG) was subtracted from the value of the gait features in the remaining gait types (L_1 , L_2 and L_3) to obtain the gait features of each gait type with respect to normal gait. Fig. 5.15 shows, in the form of bar charts, the results obtained for treadmill and overground gait trials.

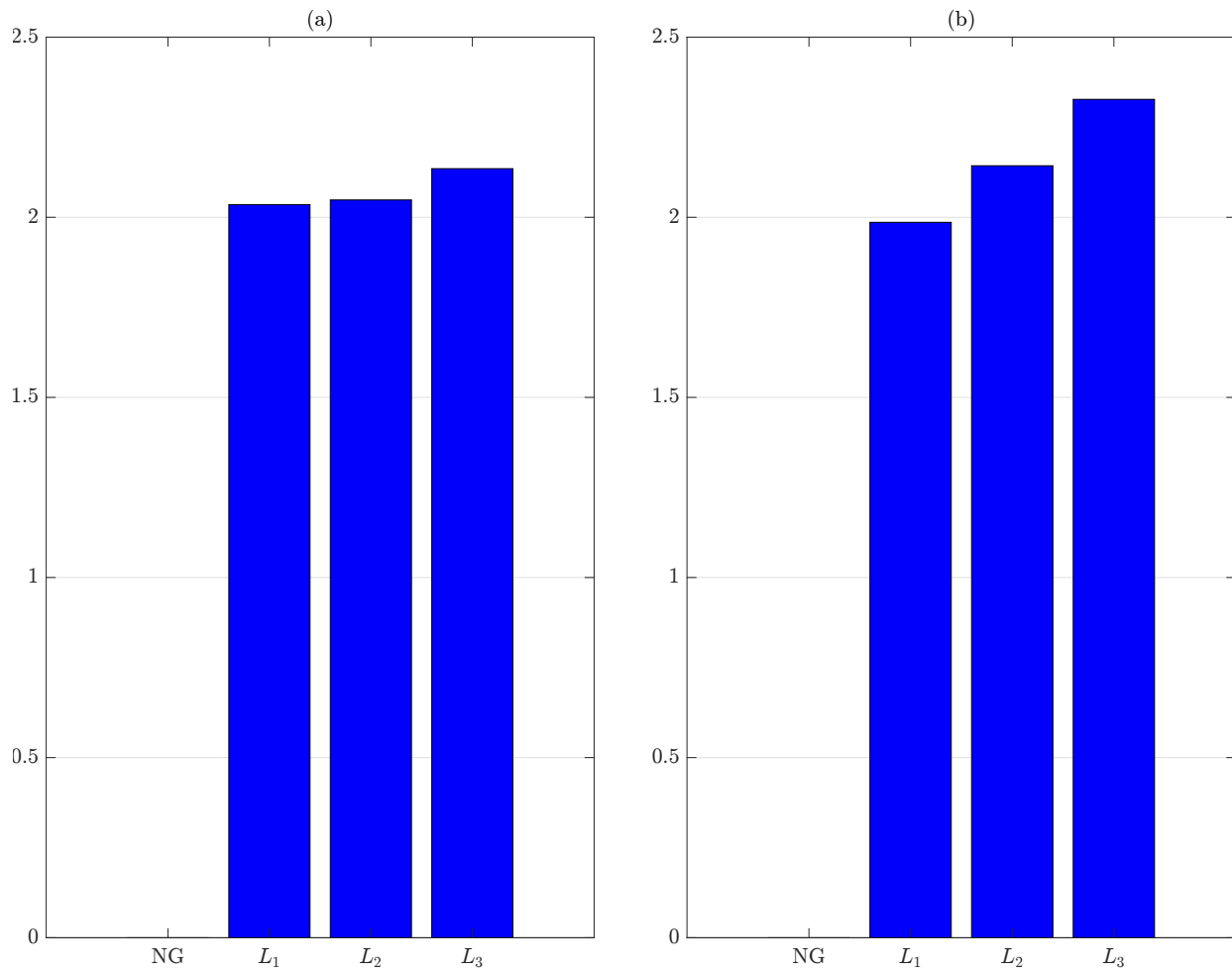


Fig. 5.15: Single gait feature of the different level of gait difficulties extracted from the data collected during (a) treadmill gait trials, and (b) overground gait trials.

Having a single gait feature that defines a gait type is quite practical when interpreting the results of the sensitivity analysis of the proposed methodology. As can be seen in Fig. 5.15, the calculated gait feature increases as the level of gait difficulty increases. This trend is evident especially for the data associated with the overground gait (see Fig. 5.15(b)). In the bar charts, we can also see that even between the lowest difficulty level and the normal gait there is a difference of at least 2 points, while the difference between gait with any additional difficulty level does not exceed 0.4 points.

The self-organization method used to extract information from pairwise comparisons between time series is not an error-free method. It should be taken into account that in a three-dimensional space it is not possible to achieve the result that the estimated location for each node completely complies with the information contained in the dissimilarity matrix, except in very remote cases. That is, there will always be an error in the positioning of the nodes. However, once the algorithm converges, the generated network contains a fairly decent approximation. Although the distances between nodes are not exactly proportional to the elements of the dissimilarity matrix, the pattern of the dissimilarity matrix prevails.

From the node coordinates, the actual geometric distances of the networks were computed using the Euclidean distance between each pair of nodes and new dissimilarity matrices were assembled, which contain the actual information embodied in the networks. These matrices are computed to verify how well the dissimilarity matrices computed from the pairwise comparison of time series were assimilated by the networks. Figs. 5.16 and 5.17 show the dissimilarity matrices recalculated from the network associated with the treadmill and the overground gait trial data, respectively.

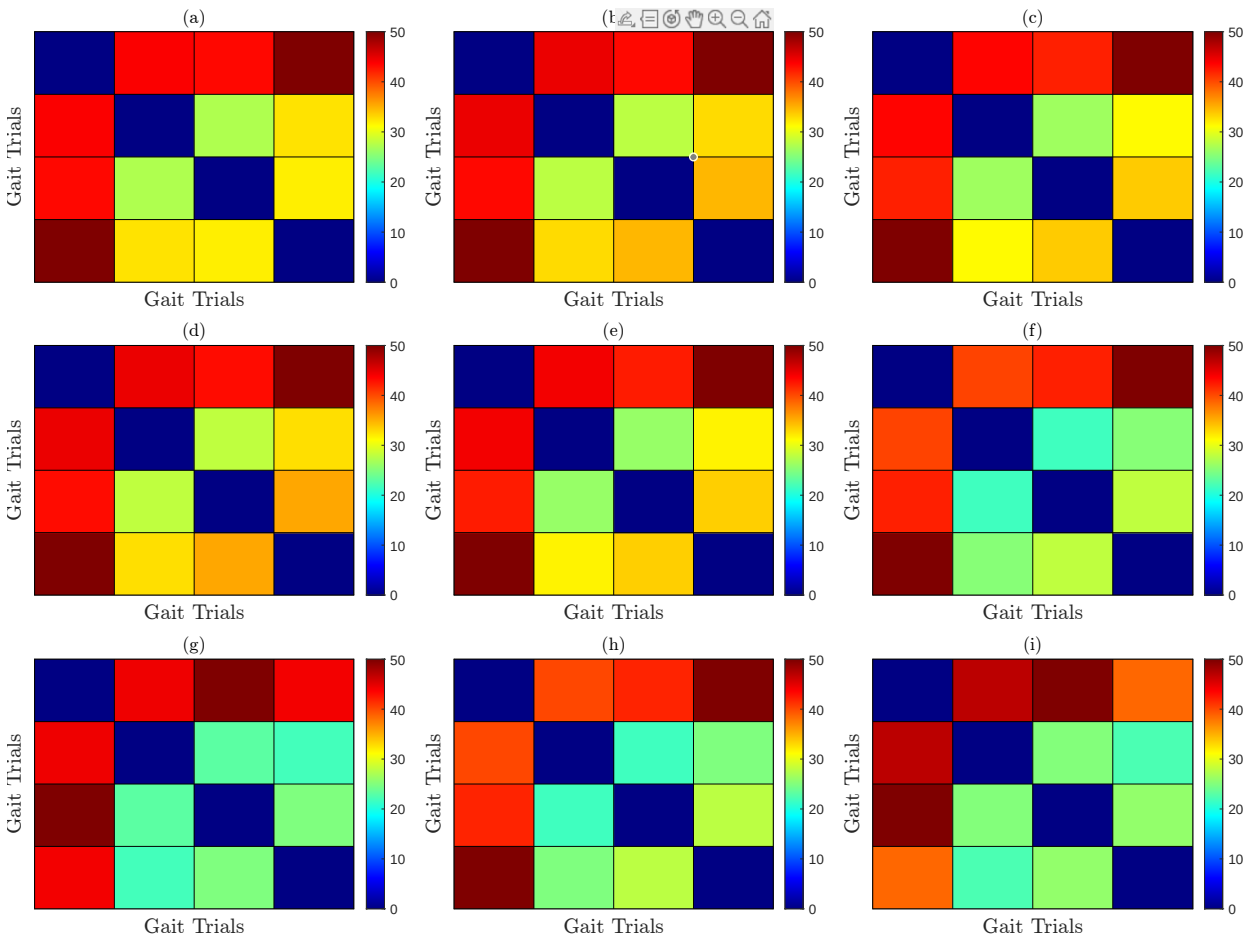


Fig. 5.16: Recalculated dissimilarity metrics based on the layouts of the network generated from the data acquired during treadmill gait trials. (a) Euclidean distance, (b) mean dissimilarity, (c) RMS dissimilarity, (d) peak dissimilarity, (e) Manhattan distance, (f) DTW, (g) inverse cross-correlation, (h) cross-correlation + DTW, and (i) Förstner distance.

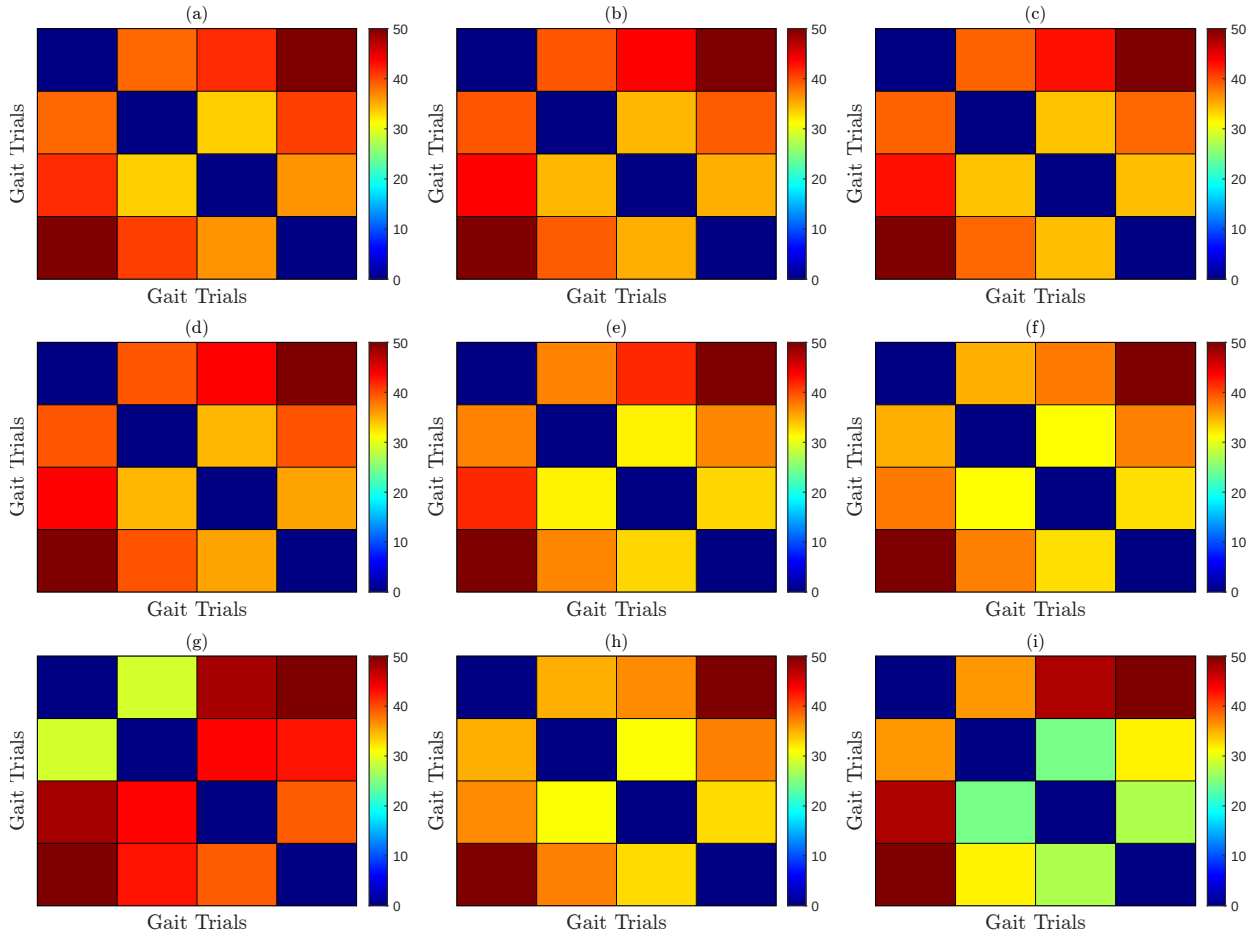


Fig. 5.17: Recalculated dissimilarity metrics based on the layouts of the network generated from the data acquired during overground gait trials. (a) Euclidean distance, (b) mean dissimilarity, (c) RMS dissimilarity, (d) peak dissimilarity, (e) Manhattan distance, (f) DTW, (g) inverse cross-correlation, (h) cross-correlation + DTW, and (i) Förstner distance.

In Figs. 5.16 and 5.18, it can be seen that the dissimilarity matrices recalculated from the generated networks, as expected, retain the same pattern as the dissimilarity matrices used to generate the networks. Unlike the original dissimilarity matrices, the color contrast of the maps is more marked. However, although the color range has changed and thus the intensity associated with similarity between elements has changed, the same clusters are observed within a new color range. The most dissimilar regions remain the same, although their intensity has changed, i.e., the first row and the first column remain the warmest colors in the map.

The new dissimilarity matrices, unlike the original ones, contain consistent information about the pairwise comparison of the multivariate series. If the self-organizing force method were used again to generate a new network, the error in node positioning would be 0, since all similarity distances contained in the matrix are geometrically achievable. From this validation stage, it can be stated that the use of the self-organizing force-driven method not only serves to visualize and reduce the data set, but also corrects the inconsistency of the dissimilarity matrices.

5.7.2 Exploring the entire dataset

In this section, we use the ICR-LLS methodology to find similar gait patterns within the entire collected dataset composed of 60 gait trials. Unlike the previous analysis, we do not analyze the subjects independently, but we determine the dissimilarity matrix containing the pairwise comparisons of the 60 multivariate series. This new analysis includes the preprocessing steps of the 60 multivariate time series (4 gait trials per subject), the pairwise comparisons between all gait trials, and the generation and self-organization of the networks.

As in sensitivity analysis, the preprocessing of the time series consists of truncating, filtering, and normalizing them. The dissimilarity matrices are then calculated using each of the proposed dissimilarity measures. Figs. 5.18 and 5.19 show the dissimilarity matrices.

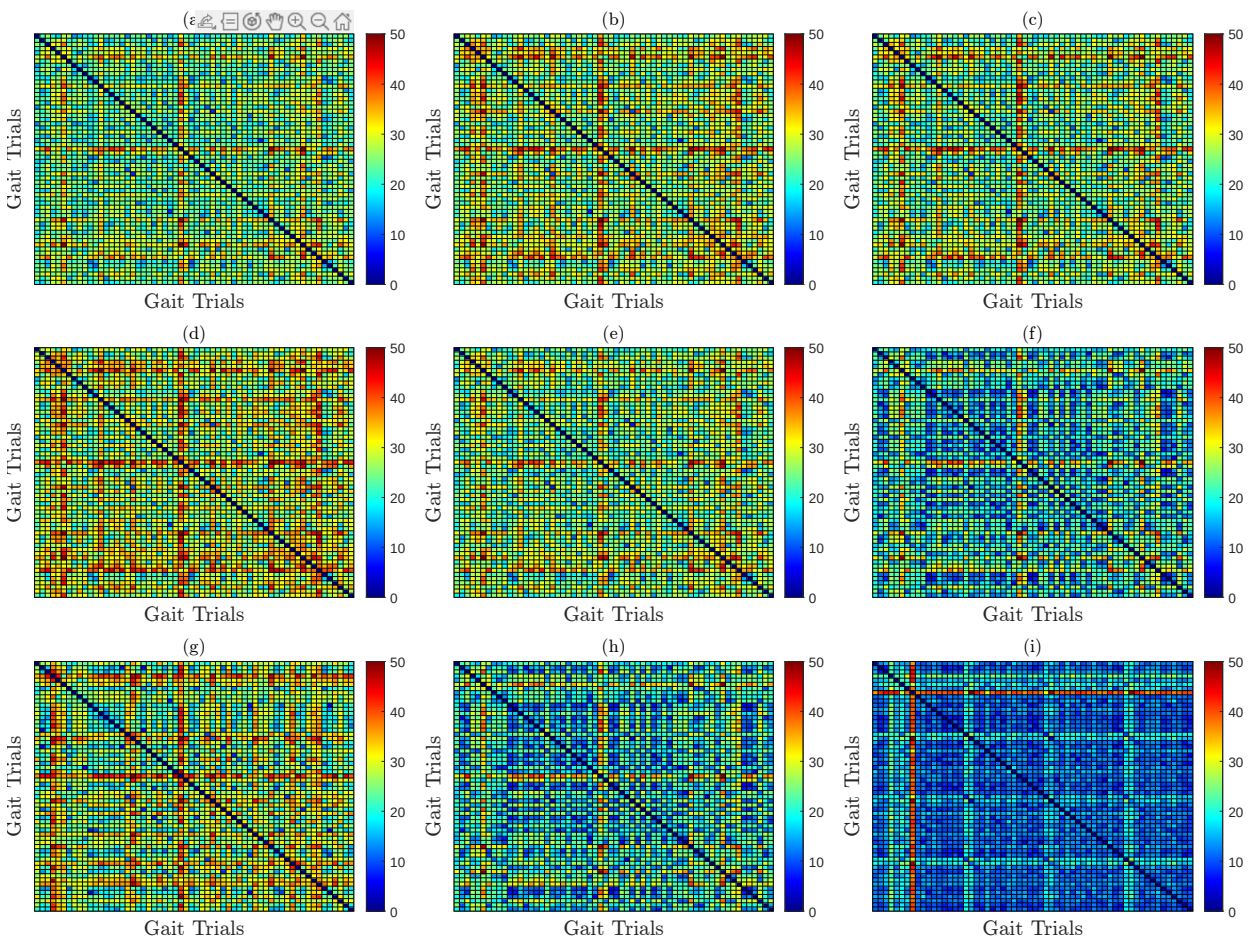


Fig. 5.18: Dissimilarity matrices calculated from pairwise comparisons of shank ICR trajectories of subjects during treadmill gait trials using (a) the Euclidean distance, (b) Mean dissimilarity, (c) RMS dissimilarity, (d) Peak dissimilarity, (e) Manhattan distance, (f) DTW, (g) Cross correlation, (h) Cross correlation + DTW, and (i) Förstner distance. In each matrix, ordered from top to bottom and from left to right, there are the gait trials corresponding to the 15 subjects with NG , L_1 , L_2 and L_3 , respectively.

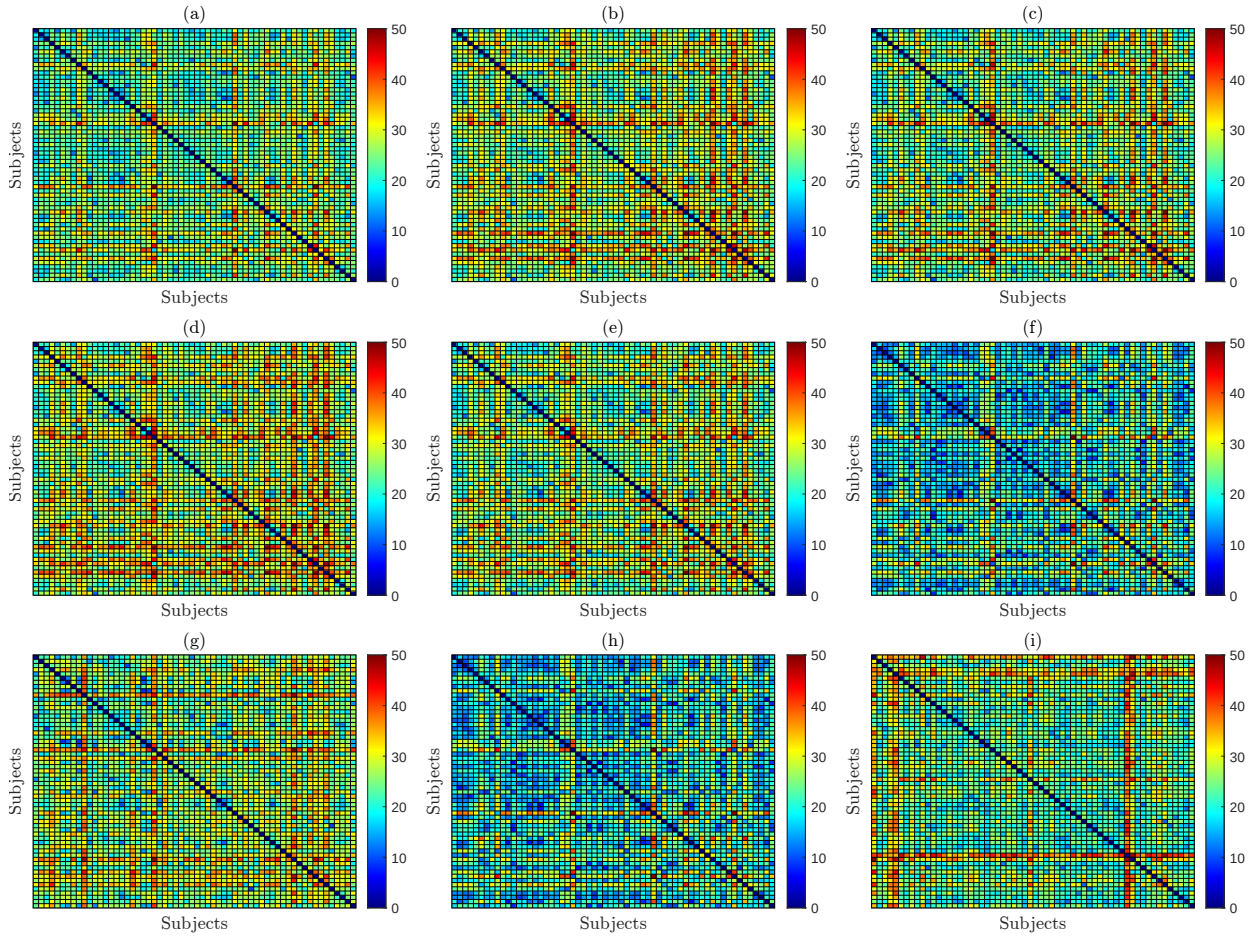


Fig. 5.19: Dissimilarity matrices calculated from pairwise comparisons of shank ICR trajectories of subjects during overground gait trials using (a) Euclidean distance, (b) Mean dissimilarity, (c) RMS dissimilarity, (d) Peak dissimilarity, (e) Manhattan distance, (f) DTW, (g) Cross-correlation, (h) Cross-correlation + DTW, and (i) Förstner distance. In each matrix, ordered from top to bottom and from left to right, there are the gait trials corresponding to the 15 subjects with NG , L_1 , L_2 and L_3 , respectively.

Due to the large number of pairwise comparisons between the gait trials that make up the dissimilarity matrices, drawing conclusions becomes challenging. Each dissimilarity matrix contains, ordered from top to bottom and from left to right, the pairwise comparisons associated with the 15 normal gait trials (NG), followed by the 15 gait trials with the 1st degree of added difficulty (L_1), followed by the 15 gait trials with the 2nd degree of added difficulty (L_2), and finally, the 15 gait trials with the 3rd degree of added difficulty (L_3). It can be seen that the dissimilarity matrix presents a different pattern when using a different dissimilarity measure despite comparing the same data. In both the dissimilarity matrices associated with the treadmill and the overground gait trials, we can notice that the dissimilarity matrix most different from the others is the matrix calculated based on the Förstner distance, shown in Figs. 5.18(i) and 5.19(i).

Focusing on the dissimilarity matrices associated with the treadmill gait trials, we can observe that in all of them there is a gait trial quite different from the rest located in the 13 row/column, indicated by the presence of a horizontal and a vertical red line. A possible explanation for this result is that the subject might feel very uncomfortable with the heel-to-sole attachment. In the matrices associated with the overground gait

trials, we observe that the most common pattern is that the comparisons with the highest intensity are located in the right zone and in the lower zone of the matrix.

The embedding of the time series in the three-dimensional network using the forced self-organization method greatly facilitates the interpretation of the information contained in the dissimilarity matrices, especially when their dimension is very large. For network generation, the algorithm was programmed to perform 8000 simulations and the initial positions of the nodes were randomly fixed.

Figs. 5.20 and 5.21 show the networks generated from treadmill and overground gait trials, respectively, using the Förstner distance-based dissimilarity matrices. The isometric, frontal, and lateral views of the three-dimensional network are presented in each figure. Each node associated with a certain level of gait difficulty is shown in a different color, so the reader can inspect where each of the groups is located, and a label indicating the subject who performed the corresponding gait trial is shown near each node.

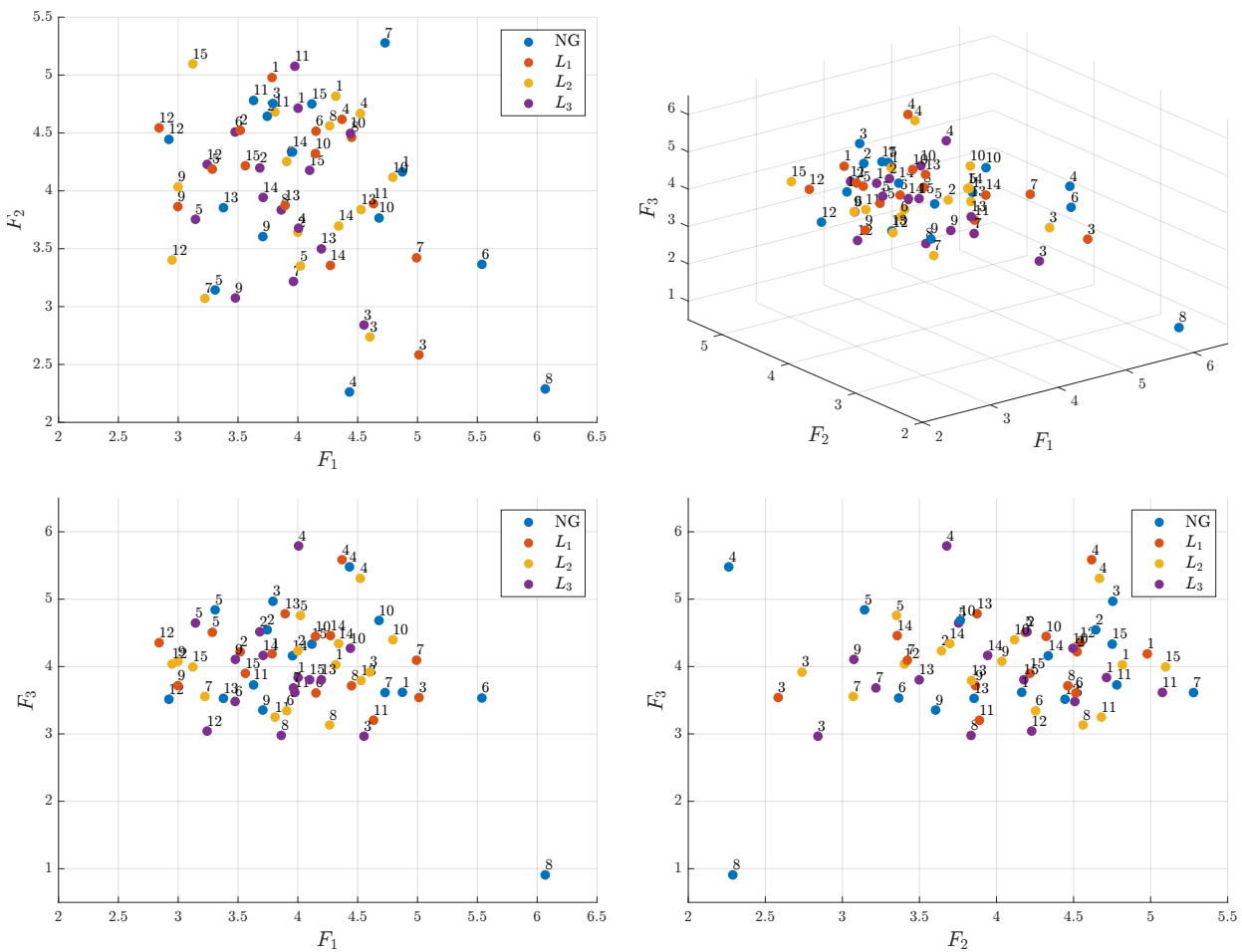


Fig. 5.20: Node arrangement generated from data collected during treadmill gait trials under different difficulty levels and using the Förstner distance for comparisons of multivariate time series. Each color represent a different level of difficulty.

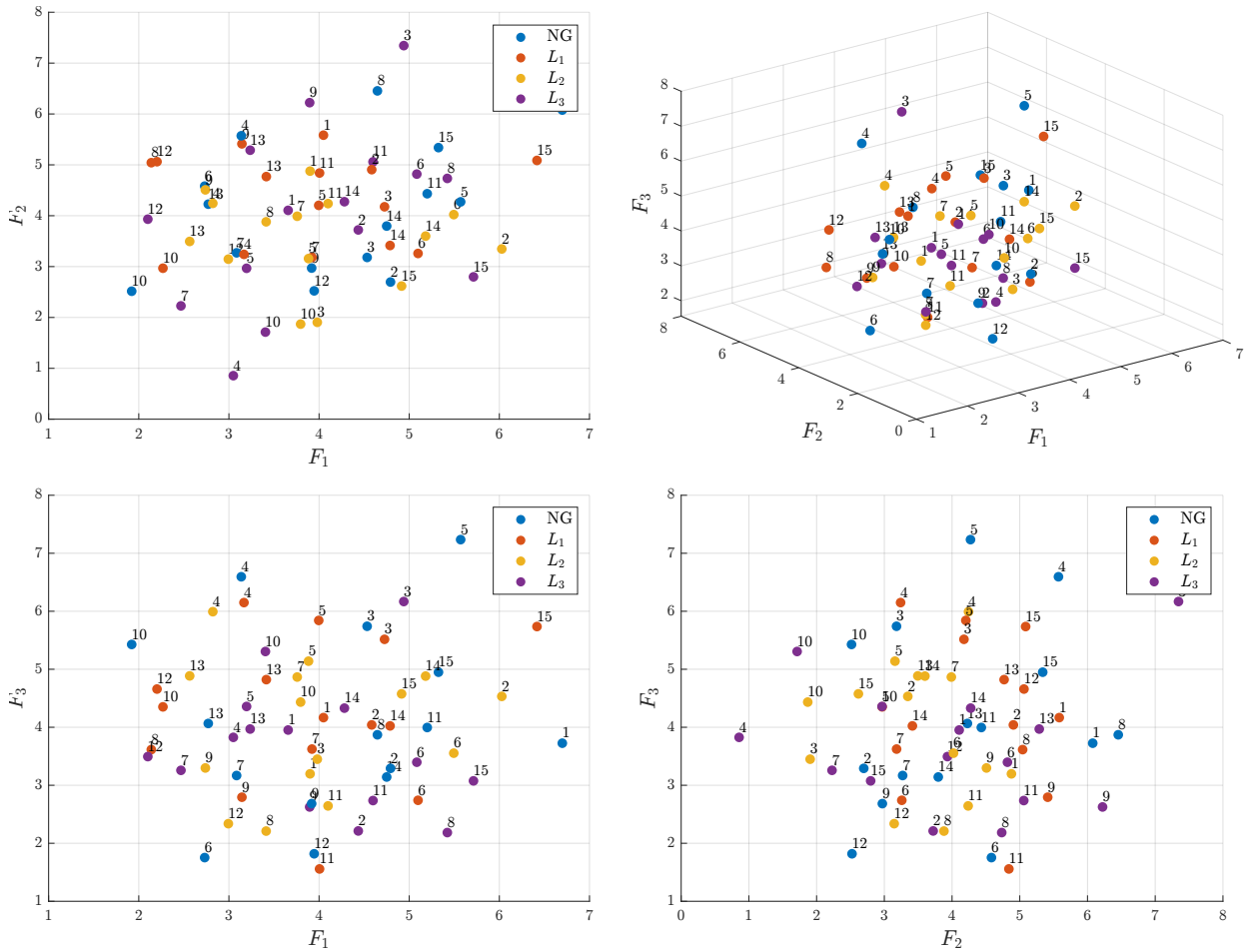


Fig. 5.21: Node arrangement generated from data collected during overground gait trials under different difficulty levels and using the Förstner distance for comparisons of multivariate time series. Each color represents a different level of difficulty.

As the reader can observe, it is hardly possible to distinguish groups of nodes in the generated network, since all nodes seem to be distributed in the same area. Only a few specific cases, such as node number 8 of *NG* (treadmill gait trials), are quite distant from the rest. The location of the nodes belonging to the groups *NG*, *L*₁, *L*₂, *L*₃ is not associated with a specific space zone in treadmill gait trials or overground gait trials. From these results, the first redeemable conclusion is that it appears that the dissimilarity between gait trials with different levels of added difficulty is not large enough for gait trials with the same level of difficulty to cluster together. These results are a setback for the proposed methodology, as the suggested hypothesis was the possibility of distinguishing different gait patterns by forming clusters in the generated network.

From a more detailed analysis of the generated networks, a second finding was reached. In most cases, the nodes corresponding to the gait trials performed by the same subject are located in the network at close distances from each other. This interesting finding indicates that indeed the methodology can recognize similar gait patterns and discriminate them; however, in the gait trials performed, despite trying to modify the normal gait pattern of the participants, it seems that it was not enough, since two gait patterns, under the same aggregate difficulty level, belonging to different subjects remain more dissimilar than two gait patterns belonging to the same subject despite adding different difficulty levels.

To help the reader more easily identify the formation of the small groups made up of gait trials performed by the same subject, Figs. 5.22 and 5.23 once again show the previously generated networks presented in x and y , but this time each gait trial performed by the same subject is shown in the same color. The labels S_1 to S_{15} indicate the subject who performed the gait trial.

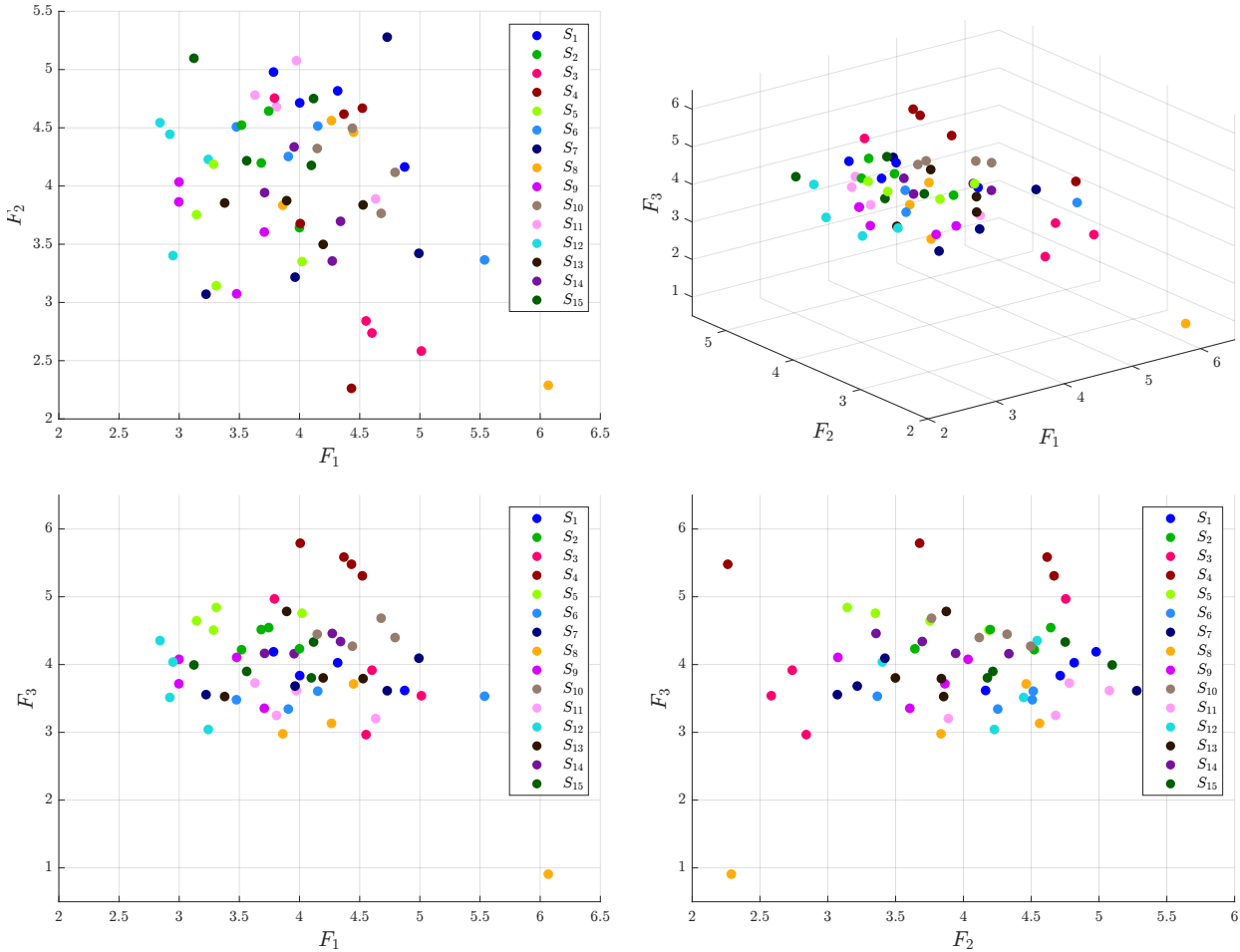


Fig. 5.22: Node arrangement generated from data collected during treadmill gait trials under different difficulty levels and using the Förstner distance for comparisons of multivariate time series. Each color represents a different subject performing the gait trials.

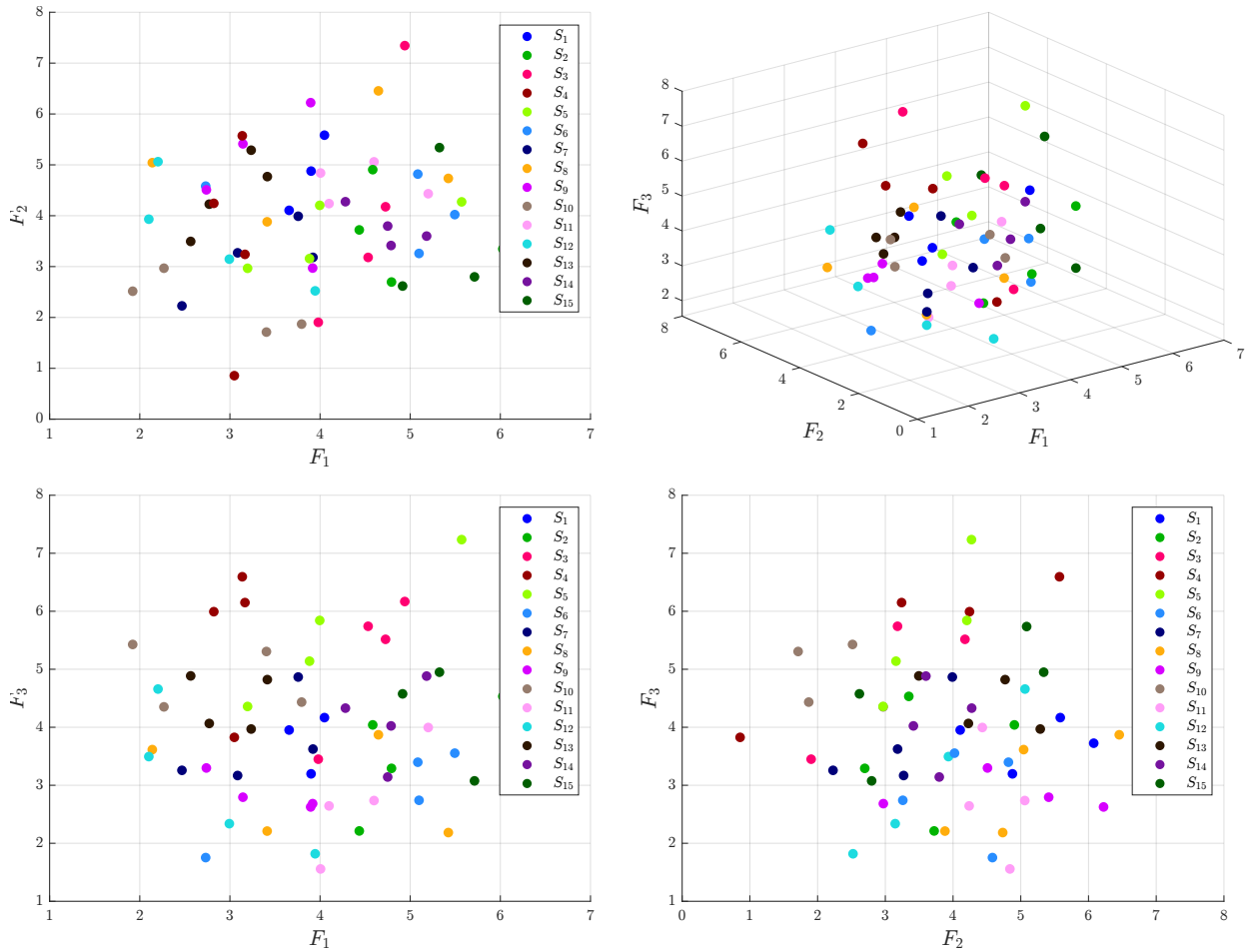


Fig. 5.23: Node arrangement generated from data collected during overground gait trials under different difficulty levels and using the Förstner distance for comparisons of multivariate time series. Each color represents a different subject performing the gait trials.

These results, although not as expected since the clusters were not formed from gait trials performed at the same degree of difficulty but by the same subject, demonstrate that the proposed methodology is capable of extracting relationships between the set of time series that are difficult to observe by simply inspecting their characteristics over time. In the gait trial performed, the gait of the subjects was not sufficiently altered to deviate from normal gait, and therefore instead of forming clusters for each type of simulated abnormal gait, small clusters were formed, each identifying the same subject.

Chapter 6

Applications of ICR-LLS to discriminate individuals with Parkinson Disease

6.1 Introduction

Parkinson Disease (PD) is a progressive chronic neurological disorder that affects the brain by destroying its parts that control movement. It is characterized by a series of motor symptoms, such as tremors, rigidity, slowness of movement, and postural instability. Furthermore, as a consequence of the progression of the disease, patients with PD in the early stages of the disease often experience impaired postural control and gait dysfunction ([Marsden, 1994](#)).

Since all forms of PD involve an alteration of the motor system, an appropriate approach to evaluate the progression of the disease over time is to assess motor symptoms. Such assessments are carried out mainly using question forms such as the Hoehn & Yahr scale, and the Unified Parkinson's Disease Rating Scale (UPDRS). The Hoehn & Yahr scale assesses the patient's ability to perform typical activities of daily living independently, with scores ranging from *I* to *V* ([Hoehn and Yahr, 1998](#)). Likewise, the UPDRS, which is the most widely used in both research and clinical settings, consists of four parts, including the assessment of intellectual status and mood disorders, the assessment of activities of daily living, the assessment of motor function, and the assessment of treatment complications. In the UPDRS, each question can receive a score from 0 (no symptoms) to 4 points (significant symptoms), and the total score is the sum of each question ([Disease, 2003](#)). A higher UPDRS indicates a more advanced stage of the disease.

Focusing on walking, gait disturbances in PD are manifested by a distinctive pattern known as "parkinsonian gait", which is an unstable movement pattern characterized mainly by walking bent forward in a stooped position, decreased arm swing, and shuffling gait, among others ([Guo et al., 2022](#)). Through the use of gait analysis, the changes produced by PD in the spatiotemporal parameters of gait, such as decreased stride length, reduced gait velocity, and difficulties in starting and stopping movements, have been identified and reported in several studies ([Kim et al., 2018](#)).

For patients with PD, the consequences of gait disturbances go beyond simple functional limitations, since in cases of severe gait disturbances, patients are more likely to experience falls. Studies on this topic have reported that 70% of patients with PD suffer at least one fall per year and 39% fall recurrently. Further-

more, median survival in patients with recurrent falls is 6 years (Allen et al., 2013). These falls contribute to fractures, injuries, fear of future falls, increased caregiver burden, hospital admissions, and ultimately decreased functional independence, highlighting the need for effective interventions.

In this section, the ICR-LLS methodology, proposed in Chapter 5, is used to discriminate gait patterns between individuals with PD and control (healthy) subjects. For this task, public data sets containing kinematic data collected from overground gait trials of healthy individuals and patients with diagnosed PD were used.

6.2 Selected kinematic gait data sets

6.2.1 Gait data set on patients with PD

Two public data sets were selected for analysis in this study. The first data set was the public data set on full-body gait kinematics and kinetics in individuals with Parkinson's disease, published by the Laboratory of Biomechanics and Motor Control of the Federal University of ABC of Brazil (Shida et al., 2023a). This data set contains data on the kinematics of the upper extremities, trunk, lower extremities, and pelvis of individuals with PD with and without medication. Kinematic data contained in this dataset were collected using a MoCap system comprising 12 Raptor-4 cameras from Motion Analysis Corporation. Furthermore, five embedded force platforms were used to collect kinetic data, including three Optima models (40×60 cm) from AMTI and two 9281EA models (40×60 cm) from Kistler. Kinematic data were acquired at 150 Hz and ground reaction force data at 300 Hz.

Raw and processed kinetic and kinematic data are provided in C3D and ASCII files, allowing for easy manipulation in advanced motion analysis programs and computing environments, such as Matlab. In addition to the files containing gait-related data for each patient, the data set provides a metadata file that contains demographic, anthropometric, and clinical data. Only files containing at least one complete GC (stance and swing phase) were considered in the analysis.

In the study, 26 individuals with PD underwent gait trials, of whom 20 were men and 6 were women. They ranged in age from 44 to 81 years, body masses from 53.3 to 94.6 kg, heights from 151.5 to 179.0 cm, and BMI from 19.2 to 34.3 kg/m². The inclusion criteria were the absence of neurological or physical dysfunction other than those associated with PD and the absence of self-reported vestibular, visual, or somatosensory dysfunction. Idiopathic individuals with PD participated in two experimental sessions for 1 week, one in the ON condition of medication and the other in the OFF condition. To be considered in the ON condition, participants had taken dopaminergic medication 1 h before the start of the session to ensure dose stabilization. In the OFF condition, the participants spent at least 12 h without PD medication.

All gait trials were performed barefoot and participants wore comfortable shorts (women wore sports bras). Participants were asked to perform gait trials on the ground at their preferred speed. Forty-four reflective markers were attached to the wearer, 26 anatomical markers to the lower limb, and 18 to the trunk and upper limb following the recommendations of the International Society of Biomechanics.

The experimental protocol followed for data acquisition comprised the following steps:

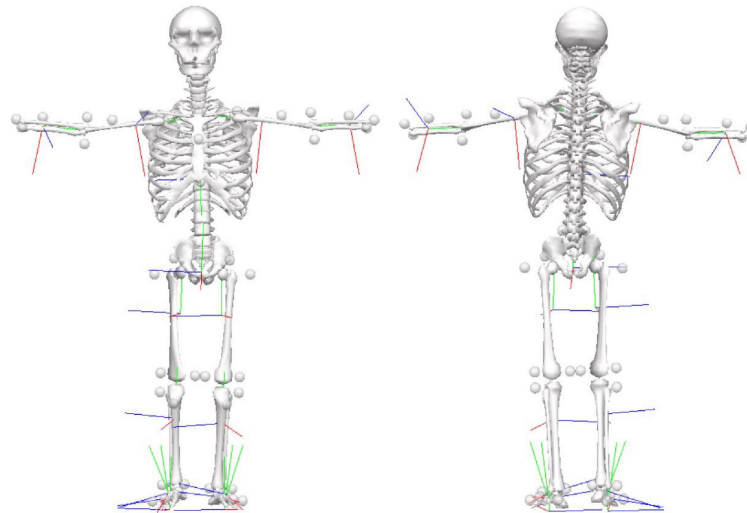


Fig. 6.1: Reference frames of the anatomical markers on the PD patients.

1. The investigator explained the data collection process to each patient and the patient signed an informed consent form.
2. The investigator interviewed the participants to collect information about their clinical data, medications, and disease diagnosis.
3. Two physical therapists with experience in movement disorders applied several scales, such as UPDRS-II, UPDRS-III, Hoehn & Yahr, and New Freezing of Gait Questionnaire (NFOG-Q), among others.
4. After the patients rested for 10 min, reflective markers were placed directly on the skin of the participant (see Fig. 6.1).
5. A calibration stage was performed with the participant motionless for 1 s, adopting a stationary T-pose with the feet oriented parallel to the x axis of the laboratory coordinate system.
6. After the calibration stage and a period of familiarization with the system, the participants were instructed to walk at a preferred speed along a 20-m walkway. Each participant performed 20 over-ground gait trials.

6.2.2 Gait data set on healthy individuals

Symptoms of PD usually appear after the age of 50 years, and its prevalence increases significantly beyond this age threshold. This trend is supported by several studies, including a community survey conducted in Aberdeen involving 100,000 residents. The survey reported that, among the participants affected by PD, there were 47 individuals aged 40 to 49 years, 78 aged 50 to 59 years, 254 aged 60 to 69 years and 832 aged 70 to 79 years (Mutch et al., 1986). Studies such as this one provide empirical evidence that PD predominantly affects individuals in advanced stages of life.

Since the mean age of the group of subjects undergoing gait trials performed in the Biomechanics laboratory of the AGH University of Science and Technology differs significantly from the mean age of prevalence

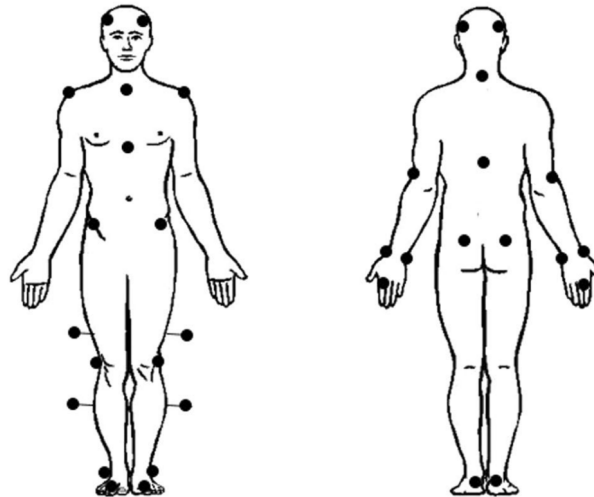


Fig. 6.2: Reference frames of the anatomical markers on the healthy individuals.

of PD, in addition to the kinematic gait data set of patients with PD, it was necessary to select a second kinematic gait data set of healthy subjects with a mean age close to 60 years. The selected gait data set was created and published by the Multidisciplinary Motor Center Antwerp movement analysis laboratory and comprises kinematic, kinetic and EMG data from 138 healthy individuals (21 to 86 years) and 50 stroke survivors (19 to 85 years) (Van Crielinge et al., 2023). From this dataset, healthy individuals were used as control subjects in the study proposed by the author of this thesis, while gait data collected from stroke survivors were not taken into account.

In recruiting healthy subjects who participated in the creation of the data set, individuals with visual impairments, abnormal gait patterns, lower extremity mobility problems, or any neurological or orthopedic disorders that could potentially affect motor performance and balance were excluded. The instrumentation used for data collection included a three-dimensional MoCap system, comprising eight Vicon T10 cameras operating at 100 frames per second with a resolution of 1 Megapixel (1120×896). Furthermore, four ground-embedded force plates were used, consisting of three AMTI type OR 6–7 force plates and one AccuGait force plate, all operating at 1000 frames per second. Raw and processed kinetic and kinematic data are provided in C3D format.

The experimental protocol followed for data acquisition comprised the following steps:

1. The investigator explained the data collection process to each patient and the patient signed an informed consent form.
2. The investigator collected anthropological data from each patient, including body mass, height, and leg length.
3. Reflective markers were attached to the bony anatomic landmarks according to the standard Plug-In Gait full-body model (see Fig. 6.2).
4. Before data collection, a static calibration was performed in which the knee flexion axis was defined and the visibility of the markers was checked.

5. Following the calibration phase and a period of familiarization with the system, participants were instructed to walk at a self-selected speed along a 12-meter walkway. Each participant performed at least 3 walking overground trials.

6.3 Preparation of dataset for analysis

From each data set, 15 subjects were randomly selected without discrimination of gender. The only criterion to be considered when selecting the subjects was that the mean age of the two groups (healthy subjects (65.3 ± 8.93) and those with PD (63 ± 10.41)) should be similar to ensure the validity of the findings. The demographic information along with the scores given by the URSD and Hoehn & Yahr scales for the case of individuals with PD is presented in tab. 6.1. The information of healthy individuals is placed in the top fifteen positions of the Table

Table 6.1: Information of the selected patients with PD and healthy individuals.

Subject number	Age	Gender	Height (cm)	Weight (kg)	UPDRS-II	UPDRS-II (Walking)	UPDRS-III	UPDRS-III (Walking)	Hoehn & Yahr
Subject 1	82	F	160	72	—	—	—	—	—
Subject 2	80	F	154	70	—	—	—	—	—
Subject 3	75	M	166.5	89	—	—	—	—	—
Subject 4	72	F	160.5	69	—	—	—	—	—
Subject 5	67	M	169	74	—	—	—	—	—
Subject 6	66	F	163.5	57	—	—	—	—	—
Subject 7	65	M	175	95	—	—	—	—	—
Subject 8	64	F	158	75	—	—	—	—	—
Subject 9	59	F	151.5	62	—	—	—	—	—
Subject 10	58	M	171.5	79	—	—	—	—	—
Subject 11	57	F	158	62	—	—	—	—	—
Subject 12	71	M	172.5	83	—	—	—	—	—
Subject 13	55	M	168.5	67	—	—	—	—	—
Subject 14	53	F	165.5	72	—	—	—	—	—
Subject 15	55	F	156.5	48	—	—	—	—	—
Subject 16	58	M	168	69	13	2	27	2	3
Subject 17	53	F	170	62.5	2	1	10	0	1
Subject 18	63	M	158.5	63.8	4	1	26	0	2
Subject 19	68	M	169	68.9	10	3	48	2	3
Subject 20	77	F	151.5	60.2	6	2	30	0	2
Subject 21	46	M	171	61.5	10	2	6	1	2
Subject 22	44	F	157	53.3	11	3	16	2	2
Subject 23	57	M	173.5	57.7	5	1	24	1	2
Subject 24	60	M	179	92.4	6	1	30	2	3
Subject 25	81	M	154.5	65.6	7	1	28	2	3
Subject 26	73	M	166	86	10	2	33	1	2
Subject 27	61	F	158.5	60.5	13	3	61	2	3
Subject 28	69	M	165	76.5	1	0	17	0	2
Subject 29	74	M	179	94.5	4	1	29	1	3
Subject 30	62	M	166.5	57.4	5	1	25	1	2

The raw data provided in the C3D files were imported into the open source Mokka software, which allows us to visualize the position of the markers collected using MoCap systems (Barre and Armand, 2014). Within the software, the markers are presented as a set of points that update their position according to the sampling rate at which the data was collected. Additionally, Mokka allows for the visualization of the ground contact forces by means of a three-dimensional vector whose origin is the position of the center of pressure

(located on the sole of the foot). Data concerning the position of anatomical markers were exported as *.xmls* files and imported into Matlab for further processing.

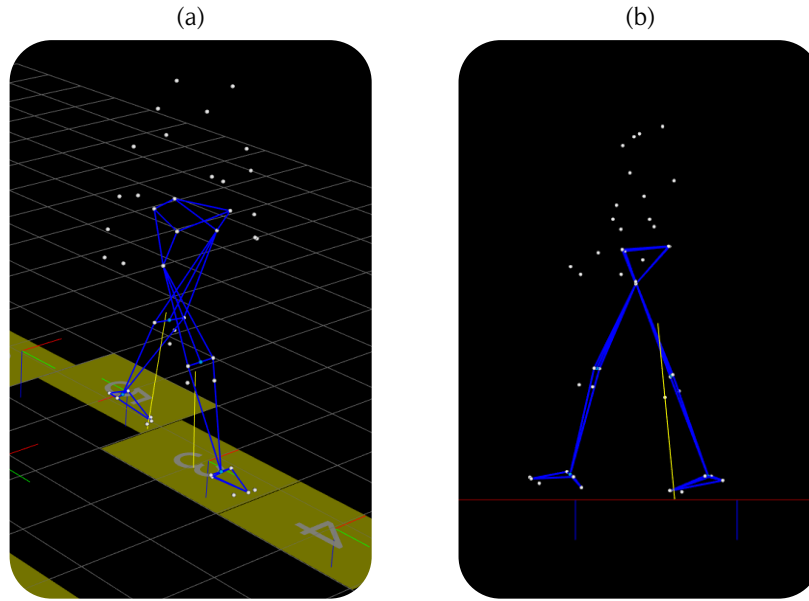


Fig. 6.3: Processing of the kinematic data contained in the selected data sets using Mokka.

The combined data set consists of the trajectory of anatomical markers of the lower extremities in the sagittal plane along the GC of 30 individuals.

Because gait data from patients with PD and healthy subjects were acquired with different motion capture systems, discrepancies would exist if the movement of the recorded anatomic markers was compared directly. Factors such as the location of the reference system from which marker displacements were measured and the fact that there may not be a direct correspondence between markers from the systems make this comparison impractical. However, as with joint motion analysis, when analyzing ICR motion (relative to the position of the center of mass of the segment under consideration), these issues are mitigated. The ICR is a kinematic parameter associated with the rotational motion of body segments, and its calculation is not significantly affected by deviations in marker location, as long as the markers remain within the same segment.

For the calculation of the shank ICR trajectories in this study, the positions and velocities of markers located at the most lateral prominence of the right lateral femoral epicondyle and the lateral prominence of the right lateral malleolus were utilized. Both gait models used in the selected datasets included these markers, facilitating the study without requiring extensive processing to establish correspondence between different models.

From the trajectory of the markers considered in the analysis, recorded in each gait trial, gait cycles were segmented, and the number of samples in each cycle was set to 200 samples by means of linear interpolation. The marker trajectories of each segmented GC were then averaged and their standard deviation calculated, resulting in three gait pattern descriptors for each gait trial. The ICR trajectories of the left and right shanks in the sagittal plane were calculated from the three descriptors, thus obtaining the distinctive average ICR trajectory ($\overline{ICR}$), its upper boundary ($\overline{ICR} + \sigma$) and its lower boundary ($\overline{ICR} - \sigma$).

6.4 Results and discussion

In Chapter 4, the distinctive trajectory of the ICR of the thigh and shank in the sagittal plane was characterized for healthy subjects in both Cartesian and polar coordinates. In this section, we extend the analysis to include gait patterns associated with Parkinsonian gait. Figs. 6.4 and 6.5 show the ICR trajectories of the thigh and shank in the sagittal plane for individuals affected by PD, in Cartesian coordinates. The corresponding decomposition of the ICR trajectories into polar coordinates is shown in Figs. 6.6 and 6.7.

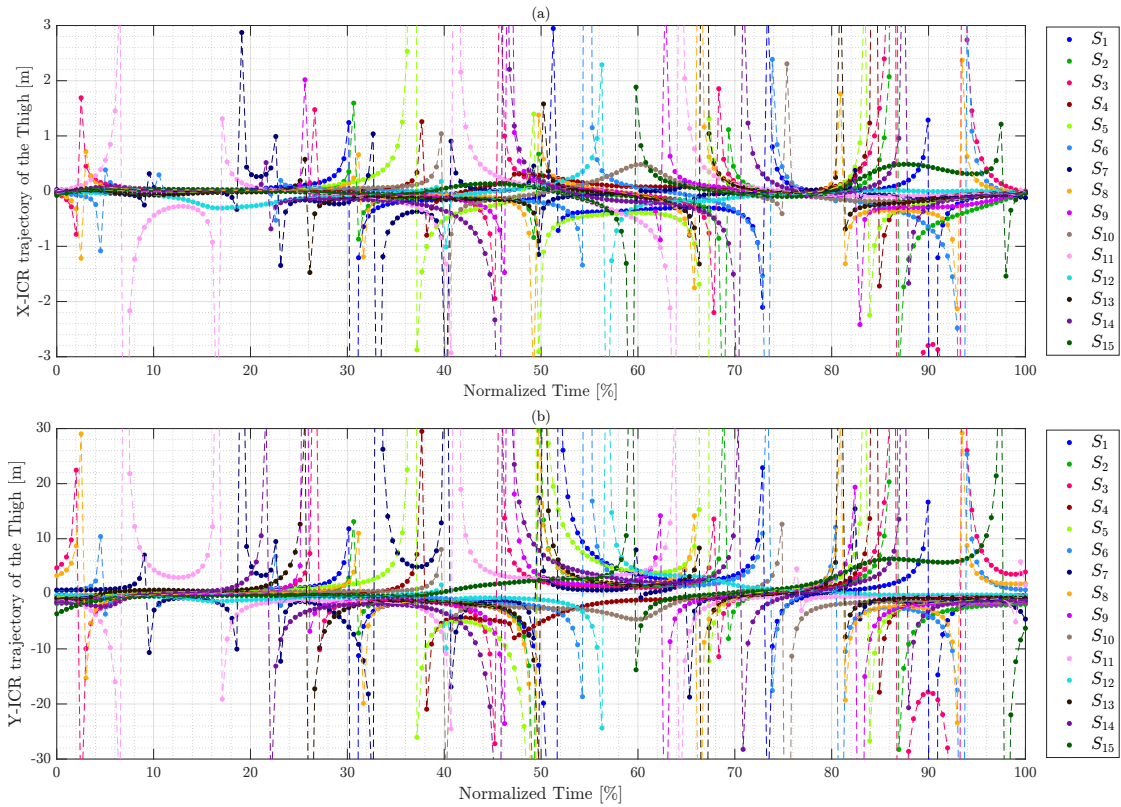


Fig. 6.4: Thigh ICR trajectories of patients with PD during overground gait trials in Cartesian coordinates. (a) Horizontal axis. (b) Vertical axis.

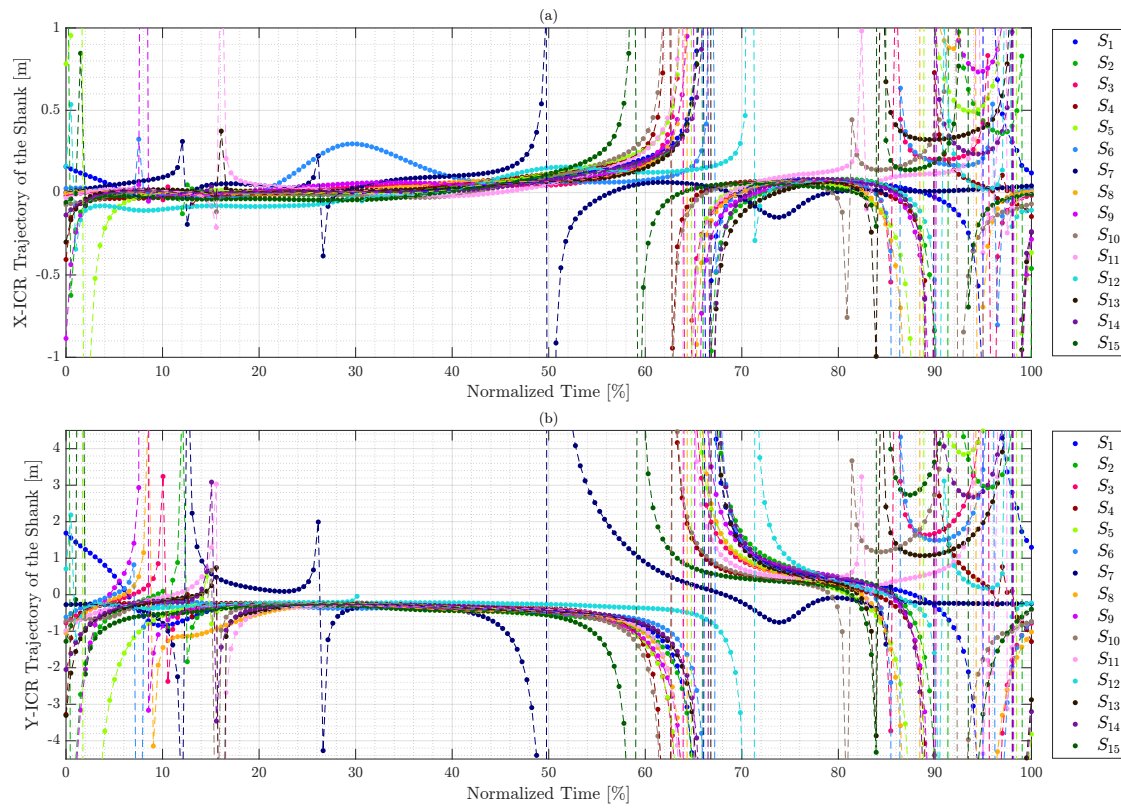


Fig. 6.5: Shank ICR trajectories of patients with PD during overground gait trials in Cartesian coordinates. (a) Horizontal axis. (b) Vertical axis.

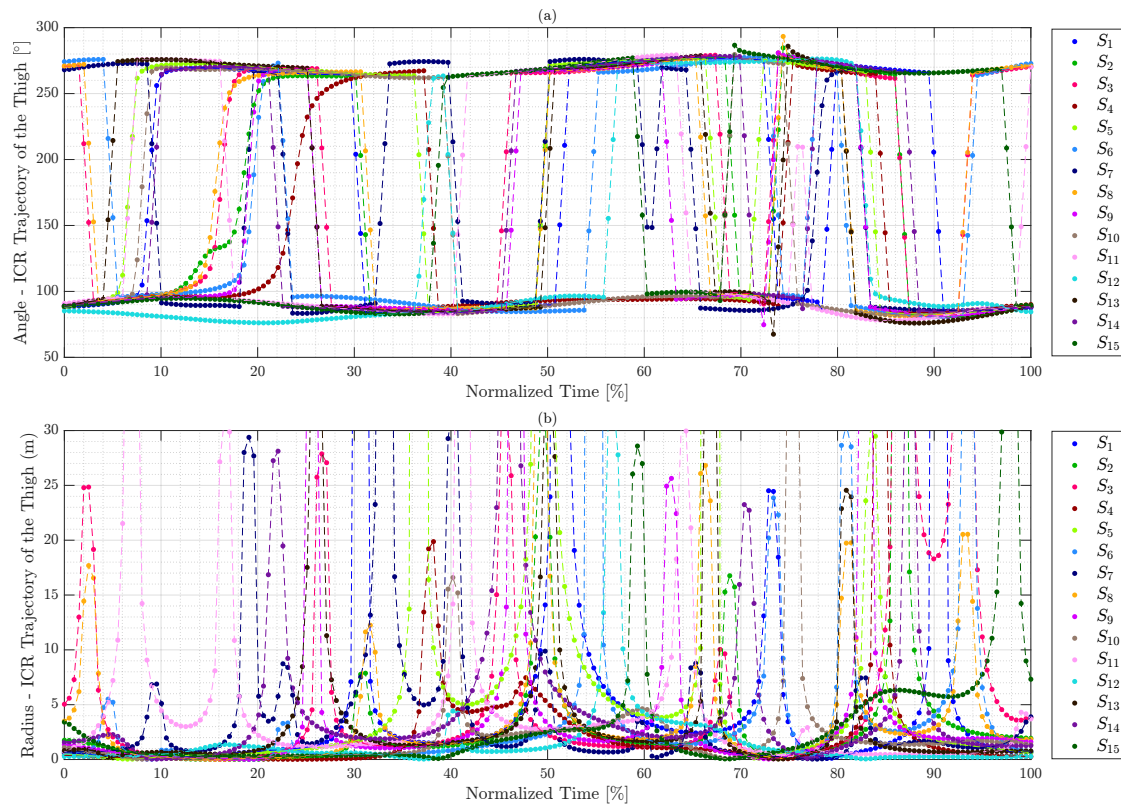


Fig. 6.6: Thigh ICR trajectories of patients with PD during overground gait trials in polar coordinates. (a) Angle with respect to the horizontal axis. (b) Radius.

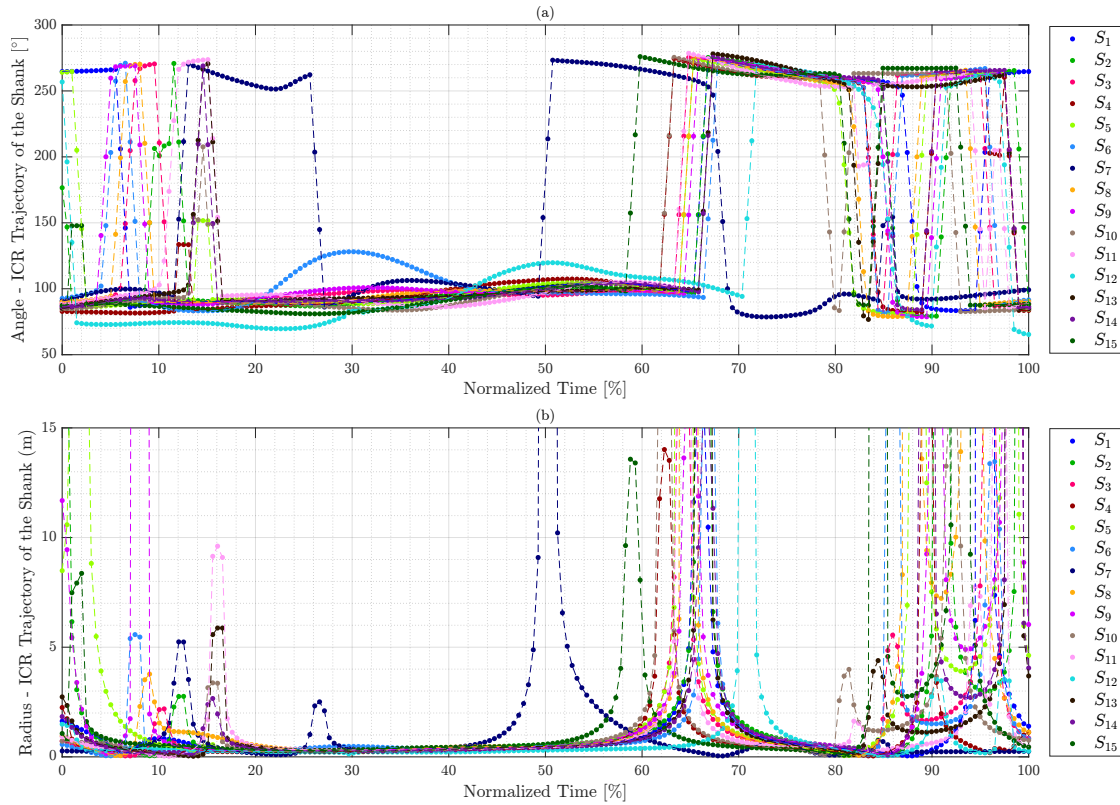


Fig. 6.7: Shank ICR trajectories of patients with PD during overground gait trials in polar coordinates. (a) Angle with respect to the horizontal axis. (b) Radius.

As the reader can see, the time series patterns corresponding to the ICR trajectory of the thigh are very different from each other, which is an expected result. Recall that the pattern corresponding to this trajectory was attempted to be characterized in Chapter 4 and it was concluded that the ICR of the thigh undergoes a chaotic motion during the GC and, therefore, is not suitable for use in gait analysis. On the other hand, the time series concerning the shank trajectory present a pattern similar to each other. When analyzing the ICR trajectories of the shank, significant variations are observed between the trajectories of the subjects, particularly during the intervals from 0% to 20% and 80% to 100% of the Gait Cycle (GC). During these periods, the trajectories exhibit asymptotic behavior. Compared to the ICR trajectories of the shank observed in healthy subjects during overground gait trials, the variability in these intervals is significantly greater. It can also be noted that the magnitudes reached by the ICR position of the shank in both x and y axes tend to be higher in those intervals, and it becomes less clear where the asymptotes occur.

Similarly, in Figs. 6.5 and 6.6, which contain the ICR trajectories of the thigh and shank in polar coordinates, we can notice the same chaotic behavior for the case of the thigh and a more defined pattern for the case of the shank.

The preprocessing of the ICR trajectories was carried out following the same procedure presented in Chapter 5. The data went through the respective stages of truncation, filtering, and normalization. Only time series concerning the ICR trajectory of the shank are considered in the application of the ICR-LLS methodology. The parameters set for each stage were the same as those set in the previous analysis.

Fig. 6.8 shows the preprocessed time series of four subjects randomly selected from the data set. Among them, two time series correspond to shank ICR trajectories of healthy individuals (N_1 and N_2), while the other two represent subjects with PD (P_1 and P_2).

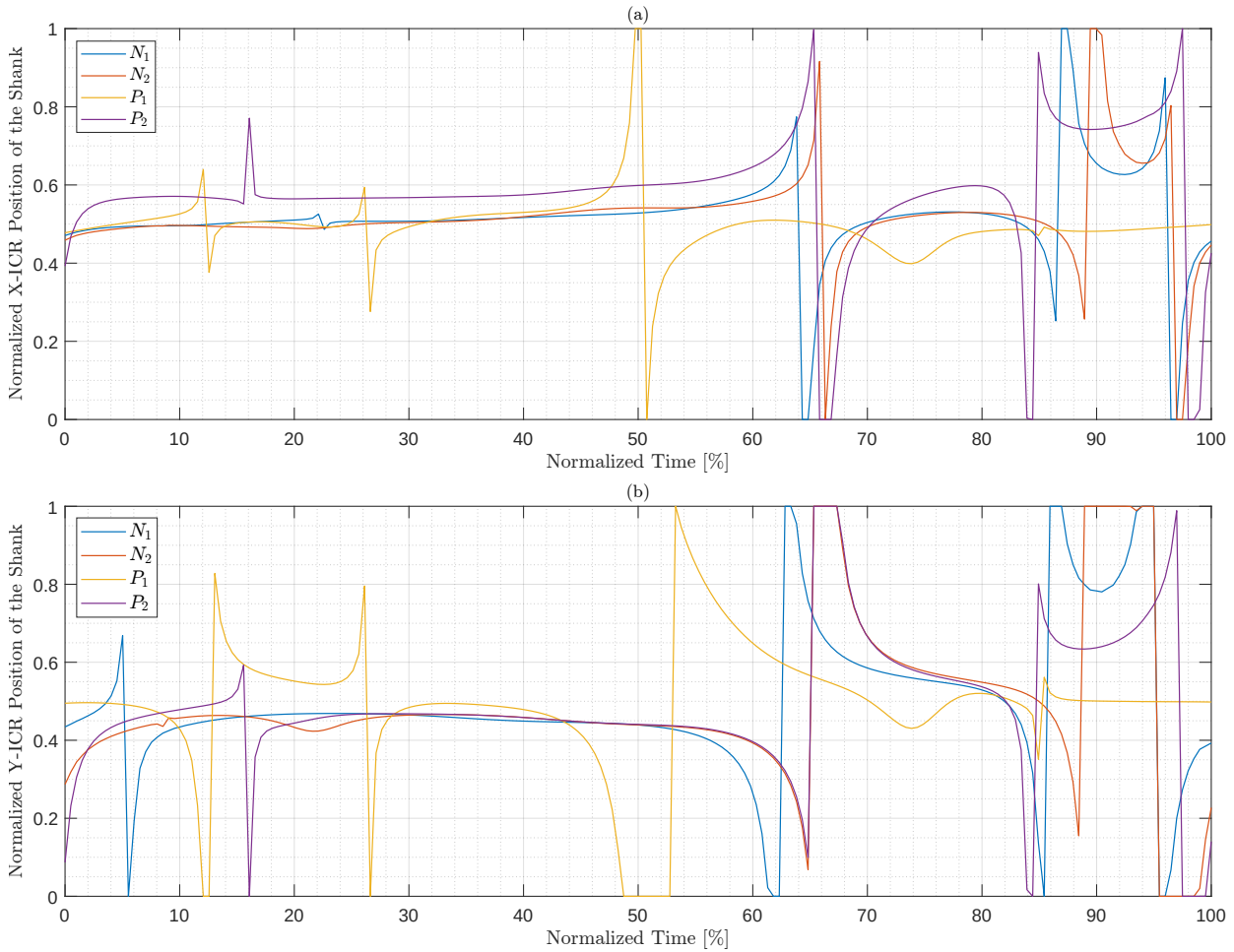


Fig. 6.8: Comparison of shank ICR trajectories of subjects with and without PD during overground gait trials. (a) Horizontal axis. (b) Vertical axis.

It can be seen that the time series pattern relative to the ICR trajectories of the shank remains similar, at least globally, despite the fact that the subject has a motor disease. An observation to highlight from these results is that more noticeable differences are perceived when comparing subject P_1 with healthy subjects, compared to subject P_2 . This suggests that accurate discrimination of subjects with Parkinson's disease from healthy individuals may not always be feasible based solely on the analyzed trajectories.

On the horizontal axis (x), it can be observed that for individual P_1 the shank ICR appears to be located more forward throughout the entire GC than for the rest of the individuals, which may be due to the posture of this individual. It is important to recall that patients with PD may undergo postural changes. On the vertical axis (y), the most significant differences are observed in the interval from 10% to 60% of the GC. While in this interval the vertical position of the shank ICR of healthy subjects remains almost constant, in subjects with PD there are spontaneous jumps.

Continuing with the next step of the ICR-LLS methodology, each gait trial performed by an individual is represented by a set of three multidimensional time series ($\overline{ICR}$, $\overline{ICR} + \alpha$, and $\overline{ICR} - \alpha$), each consisting of

the ICR trajectories along the x and y axes of the right and left shanks in the sagittal plane. The dissimilarity matrix of the data set is computed through pairwise comparisons between these time series. Here, only the results concerning the dissimilarity matrix computed using the Förstner distance dissimilarity measures are presented. As will be discussed later, the Förstner distance yields the best results among the dissimilarity measures used. The reader is referred to Appendix B.1 to check the structure of the other dissimilarity matrices.

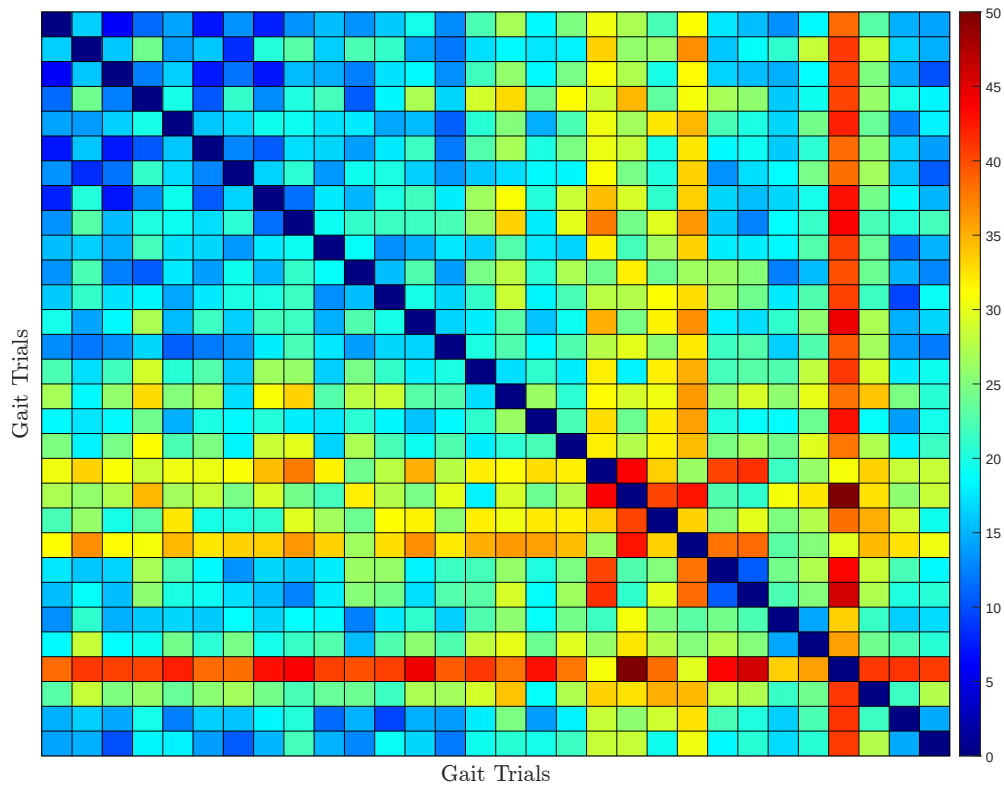


Fig. 6.9: Dissimilarity matrix calculated from pairwise comparisons of the shank ICR trajectories of subjects with and without PD using Förstner distance. In the matrix, ordered from top to bottom and from left to right, there are gait trials corresponding to 15 healthy subjects and 15 patients with PD, respectively.

The results obtained are partially consistent with the expectations raised during the formulation of the hypothesis of the study. When examining the dissimilarity matrix, four distinctive areas can be identified. In the first area (located in the upper left corner), comparisons between healthy subjects are presented. It is evident that the color map contains cool tones, suggesting a high similarity between the compared gait trials. The second and third areas (located in the lower left and upper right corners, respectively) present identical color maps. This is explained by the fact that the dissimilarity matrices are symmetric with respect to their main diagonal. Cells located in the second and third areas contain comparisons between healthy subjects and patients with PD, and, as can be seen, warmer colors compared to the first area predominate, suggesting that there are greater differences in gait patterns of these two types of gait. The last area (located in the lower right corner) shows comparisons between subjects with PD. This area appears to be the warmest of all, suggesting that there is significant variability in gait pattern among PD patients despite all having the same disease.

Using the calculated dissimilarity matrices, the described force-directed self-organization method was used to embed the multivariate time series, representing each gait trial, in a 3D space. To generate the network, a simulation consisting of 100,000 iterations was performed, resulting in the estimation of the spatial location of each node (gait trial performed by a subject). Fig. 6.10 shows the layout of the network generated from the dissimilarity matrix computed using the Förstner distance. This figure includes an isometric view and three orthogonal views to provide a comprehensive understanding of the estimated spatial distribution of the nodes.

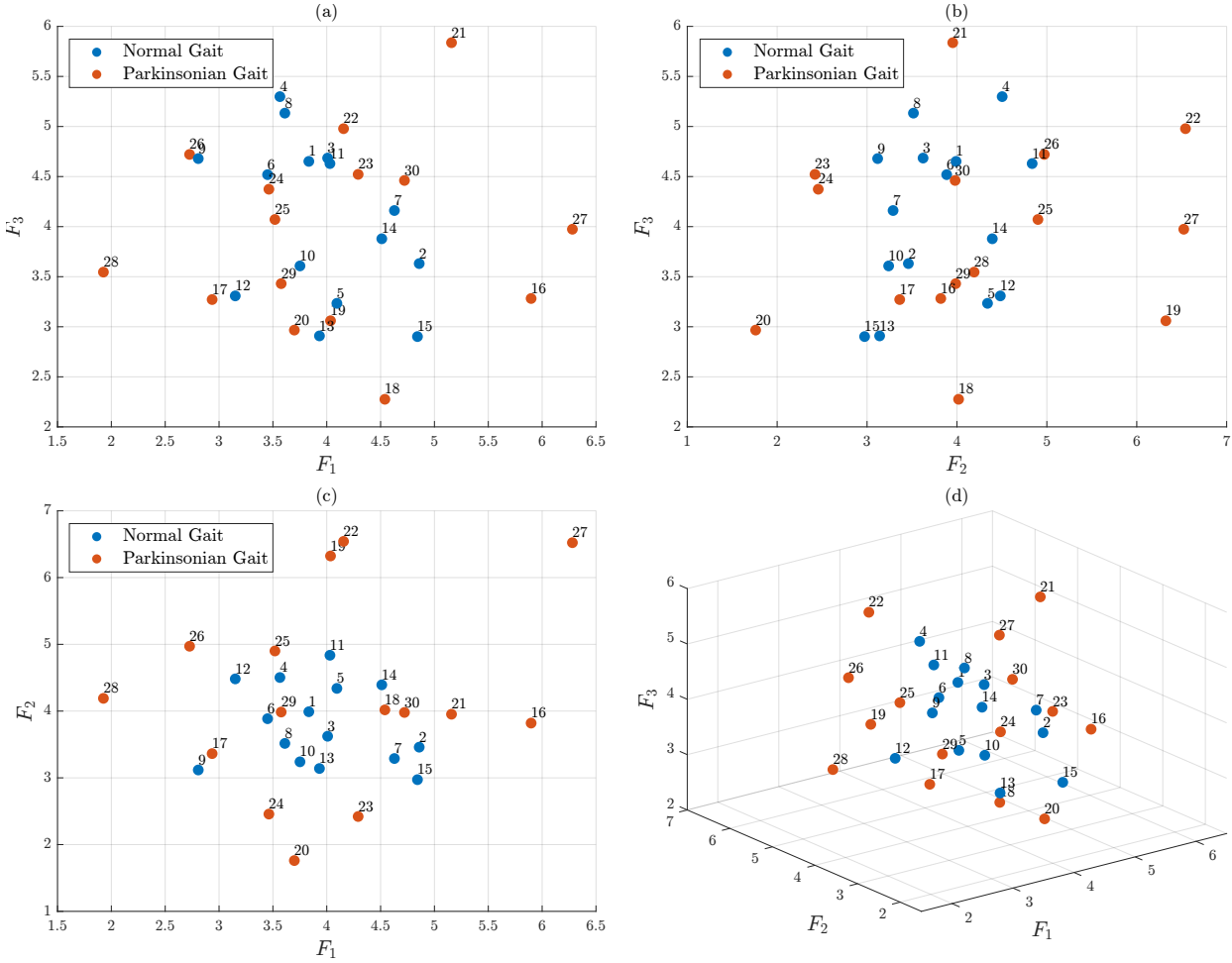


Fig. 6.10: Network layout estimated from the Förstner distance-based dissimilarity matrix. (a) Front view. (b) Lateral view. (c) Top view. (d) Isometric view.

As anticipated from examining the Förstner distance-based dissimilarity matrix, the layout of the generated networks reveals the presence of data clusters in different regions. The nodes representing the gait trials performed by healthy individuals tend to be located centrally in the network, around coordinates $[4, 4, 4]$. In the vicinity of this center, a primary cluster is distinguishable, consisting primarily of nodes corresponding to healthy individuals, although a few trials from subjects with PD are also found nearby. The nodes representing the gait trials of patients with PD are located primarily outside of this main cluster, appearing as outliers relative to the group of healthy individuals.

Visual inspection confirms that the top view of the node arrangement provides the clearest discrimination between groups. In this view, with some exceptions (nodes 29, 18, and 30), the nodes representing healthy

subjects are closely positioned, indicating low intragroup variation. Other views show a wider dispersion of these nodes, complicating the discrimination task. The isometric view further supports this observation, as the nodes representing healthy subjects are clustered near the center of the space.

The generated network provides a graphical means to understand the similarity between the gait patterns of healthy subjects and those with PD and to discriminate groups with similar patterns. The ICR-LLS methodology could end this step, as by visual inspection the clinician would be able to identify subjects whose gait patterns are slightly different from the patterns of healthy individuals and possibly diagnose a health condition. However, it is still possible to automate pattern discrimination and provide the clinician with more information to consider when making decisions. In consequence, the last stage of the ICR-LLS methodology, omitted so far, is executed to automate the recognition of clusters formed by the self-organization method. This step, although optional, provides the clinician with additional information that is useful in making decisions about whether a given subject likely has a health problem that affects his gait.

Cluster analysis of gait-related data, represented by nodes in a 3D space, was carried out using the DBSCAN algorithm. For a detailed explanation of this density-based clustering technique, refer to Section 5.6. To achieve optimal performance, multiple simulations were performed by varying the algorithm's hyperparameters, specifically the minimum number of elements in a cluster ($MinPts$) and the neighborhood radius (ϵ). The results of these simulations were evaluated using both internal and external metrics, namely the silhouette score (SC) and normalized mutual information (NMI). The score values from each simulation were then used to create a grid that served as a search tool to identify optimal parameter tuning. Fig. 6.11 illustrates the grids generated. The values of ϵ ranged from 0 to 2.5 in increments of 0.01, while the values of $MinPts$ ranged from 3 to 9.

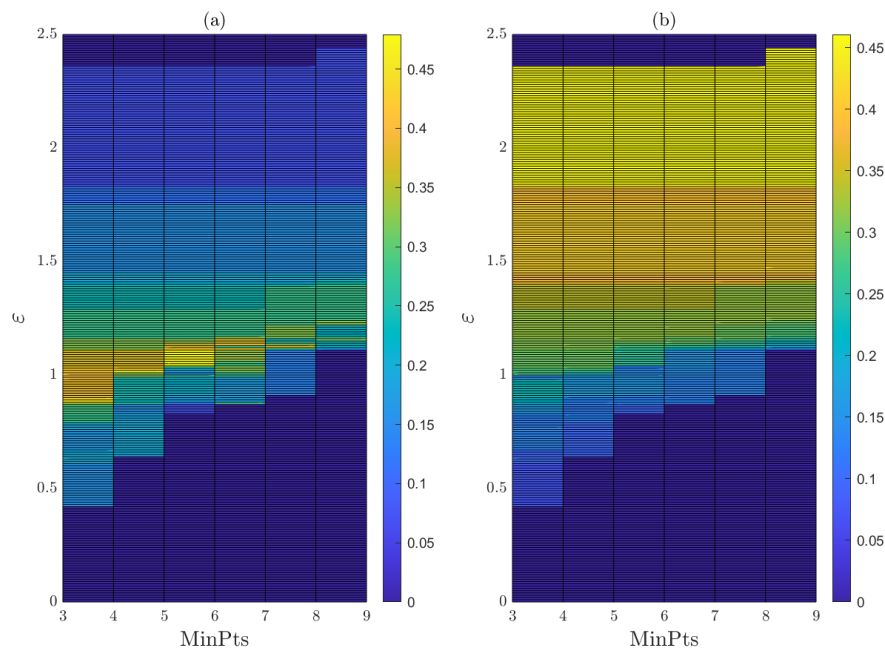


Fig. 6.11: Grid constructed to search for the optimal tuning of the clustering algorithm parameters. (a) Normalized Mutual Information (NMI). (b) Silhouette score (SC).

Fig. 6.11 shows several regions in which the optimal set of parameters can be explored. Regions where both scores have zero values should be excluded, as they correspond to parameters that give rise to a single cluster. We observe a negative correlation between both scores, as high values of one score correspond to low values of the other. This suggests that the best algorithm partition may not align consistently with the discrimination between healthy subjects and subjects with PD.

The constructed grid serves to guide the elimination of parameter combinations that lead to poor results. However, it is essential to visually inspect each solution to assess the suitability of the resulting partition. Fig. 6.12 shows a comparison between the actual distribution of healthy subjects and subjects with PD and the groups identified using the set of algorithm parameters with the best performance resulting from the search within the grid.

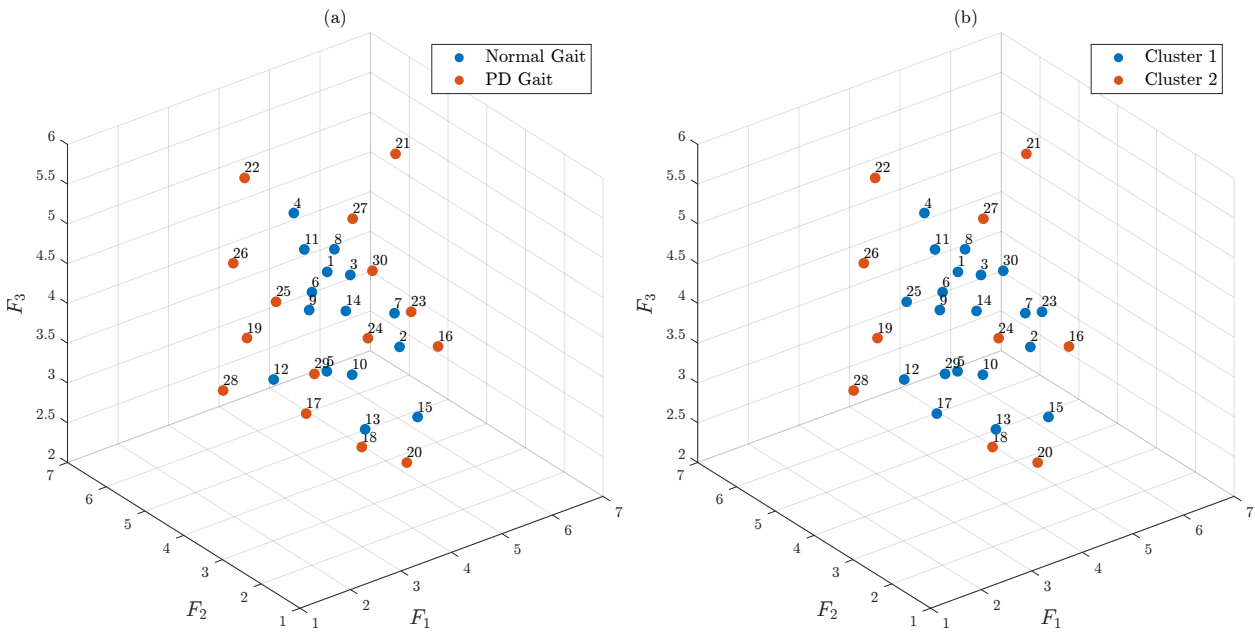


Fig. 6.12: Clusters identified from the Förstner distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.04$ and $MinPts = 5$. (a) Actual distribution of subjects. (b) Identified clusters.

The application of the DBSCAN algorithm resulted in the identification of two clusters with uneven distribution of elements. One cluster, which comprises the smallest number of elements, includes exclusively subjects diagnosed with PD, while the larger cluster contains predominantly healthy subjects. Although this result does not perfectly reflect the ground truth, it demonstrates a positive scenario in which the identified clusters show substantial correspondence with the actual desired partition.

Upon examining the cluster consisting of subjects with PD, it is evident that these nodes are relatively far apart, indicating significant intragroup variation. In addition, these nodes are also distant from elements within the other cluster, suggesting a high variation between groups. This spatial arrangement implies that the cluster that contains subjects with PD resembles more of a collection of outliers than a cohesive cluster.

The clusters identified by the DBSCAN algorithm are very sensitive to hyperparameter tuning. Adjustment of these parameters can lead to significant changes in the assignment of elements to clusters, and can also result in the identification of more than two clusters. Although diagnostic tasks, such as the one addressed in this study, often involve binary classification, the presence of multiple clusters does not neces-

sarily mean that the method is not working properly. In cluster analysis, the goal is to uncover the inherent relationships in the data without presupposing any underlying truth. The resulting partitioning is naturally derived from the distinctive features of the data, and no rules are applied to impose the desired solution. The identification of more than two clusters may suggest, for instance, the potential of the method to discern different levels of severity of PD.

To assess the performance of the partition generated by the clustering algorithm, both SC and NMI are computed. The SC score is the average silhouette value across all elements in the dataset and serves as an overall assessment of the quality of the clusters. However, a more accurate approach is to examine individual silhouette values to identify items that may loosely belong to various clusters.

Fig. 6.13 presents the SC values for each element in the data set. High positive values indicate that elements are confidently assigned to a cluster, which is especially evident in the case of elements belonging to the cluster of healthy individuals. In contrast, lower SC values or negative values suggest that items are ambiguously assigned to clusters, indicating possible noise or atypical behavior, as seen in the clusters associated with subjects with PD.

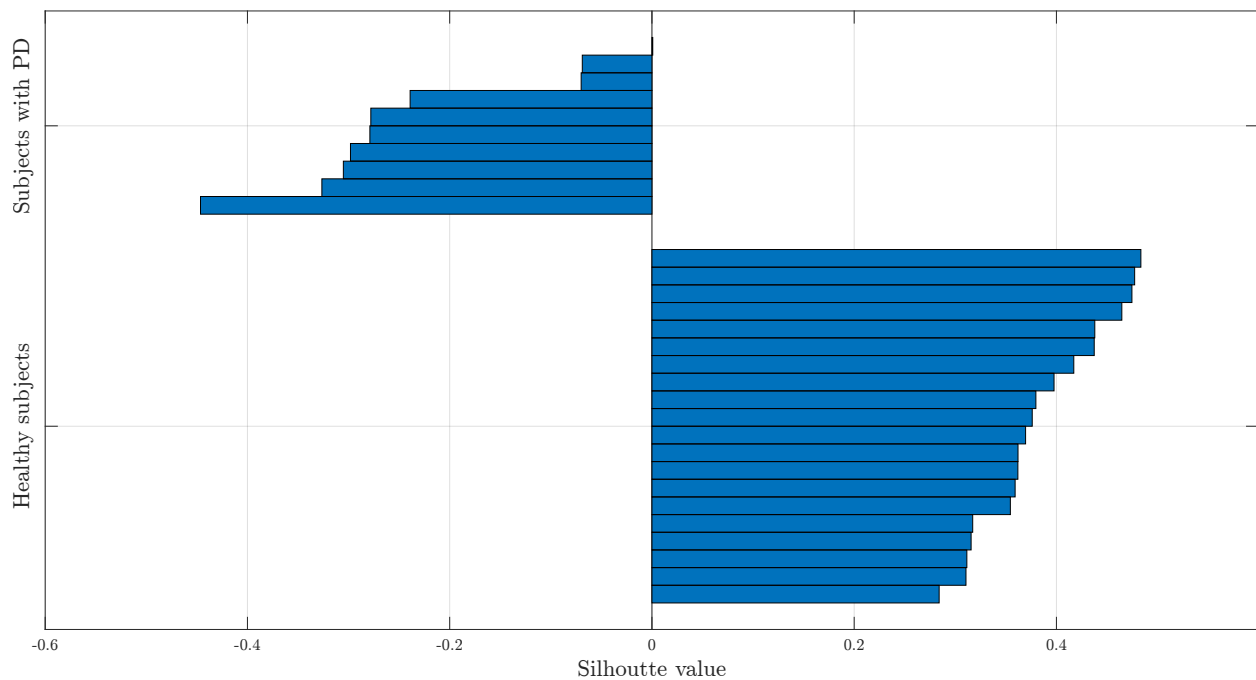


Fig. 6.13: Silhouette values for element clustering derived from the Förstner distance-based network layout using the DBSCAN algorithm with parameters $\varepsilon = 0.87$ and $MinPts = 3$.

Up to this point, only the results of applying the DBSCAN algorithm specifically to the network layout derived from the Förstner distance have been presented. However, the DBSCAN algorithm was systematically applied across all calculated dissimilarity matrices, adjusting the hyperparameters to achieve two or three clusters. For a comprehensive view of the network layouts and clusters identified using each dissimilarity matrix, see the appendix B.1.

In each application of the DBSCAN algorithm, the clustering performance was evaluated using SC and NMI scores. The performance evaluation for each of the partitions generated derived from the different dissimilarity matrices is summarized in Table 6.2.

Table 6.2: Summary of clustering performance evaluations across different dissimilarity measures and clustering algorithm settings.

	# Clusters = 2				# Cluster = 3			
	ε	MinPts	SC	NMI	ε	MinPts	SC	NMI
Euclidean Distance	1.79	7	0.163	0.197	1.61	5	0.184	0.218
Mean Dissimilarity	1.35	5	0.162	0.220	1.52	3	0.159	0.236
RMS Dissimilarity	1.47	6	0.215	0.237	1.29	3	0.127	0.281
Peak Dissimilarity	1.61	6	0.204	0.237	1.42	4	0.195	0.259
Manhattan Distance	1.59	6	0.189	0.278	1.47	6	0.166	0.224
Dynamic Time Warping (DTW)	0.92	5	0.185	0.322	0.91	4	0.188	0.245
Cross-Correlation	1.22	3	0.148	0.206	1.07	3	0.237	0.110
Cross-Correlation + DTW	0.89	4	0.212	0.370	0.85	3	0.291	0.214
Förstner Distance	1.04	5	0.256	0.479	0.99	4	0.317	0.150

Table 6.2 presents the internal and external evaluation of partitions generated using DBSCAN with various dissimilarity metrics. The results indicate that the Förstner distance produced the closest clustering to the ground truth, achieving an NMI of 0.4 with an ε value of 1.04 and a *MinPts* value of 5. This parameter configuration identified a well-defined cluster and clusters consisting of outliers. Although none of the DBSCAN runs perfectly identified the true distribution of the data, most of the misclassified items corresponded to subjects with PD, suggesting that the method avoids false positive diagnoses. To check the veracity of this statement, refer to the Appendix.

Among the dissimilarity measures, the Förstner distance consistently obtained higher NMI scores when only two clusters were identified, one of them corresponding to noise. However, the NMI score decreased significantly when three clusters were identified. This is because the ground truth (actual distribution of the data) corresponds to two clusters, making a perfect NMI score inherently impossible with three clusters. However, the detection of a third cluster may have diagnostic relevance, as it could indicate different levels of disease or similarities between the gait patterns of some healthy subjects and those with PD, suggesting a possible early stage of the disease. Fig. 6.14 shows the partition obtained using the Förstner distance when three clusters are identified.

In the calculation of the SC score, the elements of the identified cluster of outliers were excluded, as their nature usually results in negative silhouette values, as shown in Fig. 6.13. Despite this exclusion, the SC score did not exceed 0.5 in any scenario. The highest SC score obtained was 0.317 using the Förstner distance, indicating that the clusters are not strongly formed and exhibit moderate cohesion and separation. The rather weak SC score can be attributed to demographic differences between subjects, such as age, sex, and body proportions, which affect gait patterns in addition to the disease studied. For a more detailed analysis, the readers are referred to the appendices, which provide the distribution of nodes obtained from each dissimilarity measure.

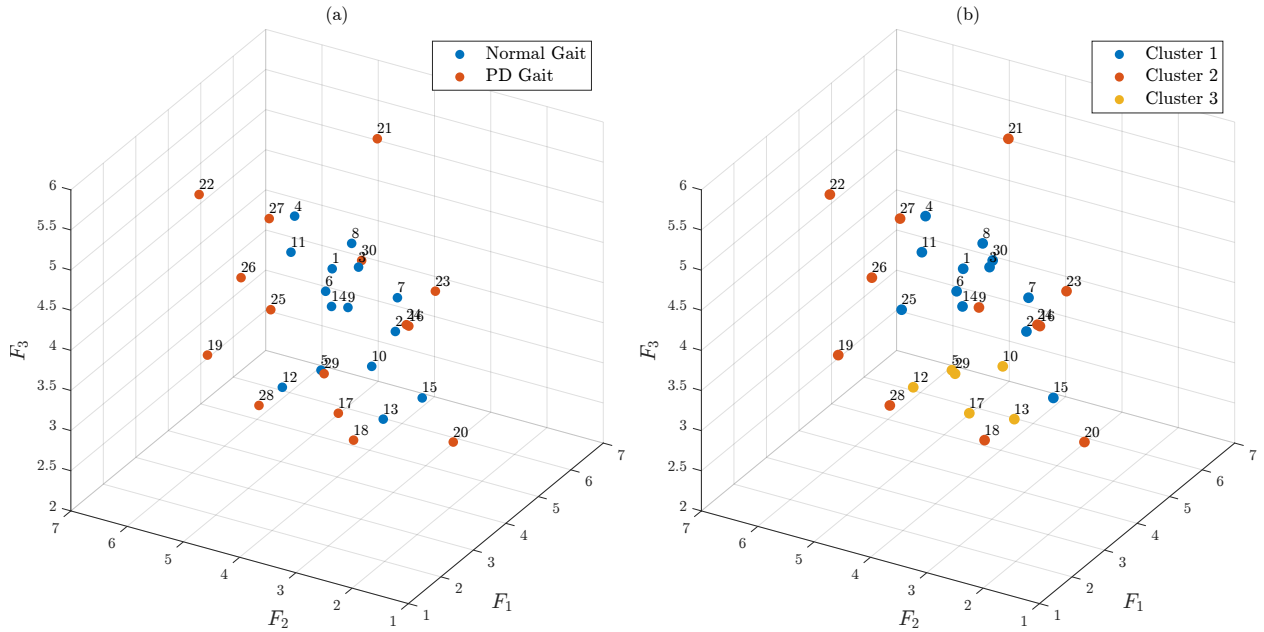


Fig. 6.14: Clusters identified from the Förstner distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 0.99$ and $MinPts = 4$. (a) Actual distribution of the subjects. (b) Identified clusters.

In applying the proposed methodology to distinguish subjects with PD in a clinical setting, the absence of prior knowledge of ground truth poses a significant challenge. Although collecting data from individuals with and without diagnosed PD provides some level of known ground truth, the actual health status of the subjects targeted for diagnosis remains unknown. Therefore, in the optimal scenario, the selection of algorithm parameters should be based primarily on internal clustering metrics, such as silhouette scores (SC). Following this idea, several simulations were performed in which the algorithm parameters were systematically varied. The process was carried out according to the information presented in Fig. 6.11(b), which presents the SC values obtained from the different settings of the parameters. Specifically, the task consisted of exploring the regions of the grid where the silhouette scores are greater than 0, while avoiding parameter configurations that produce scores close to the maximum value.

A cluster ensemble technique was used to integrate multiple partitions, obtained by varying the set of parameters of the DBSCAN algorithm, into a unified consensus clustering solution (Nguyen and Caruana, 2007). The approach followed consists of calculating a co-association matrix, where each entry in the matrix represents the frequency of each pair of elements assigned to the same cluster when the parameters of the clustering algorithm are changed. Subsequently, a secondary clustering algorithm less sensitive to hyperparameter changes, such as agglomerative clustering, is applied to derive final consensus labels based on the information contained in the co-association matrix. By aggregating information from various clustering results, this approach ensures a cohesive clustering solution that consistently captures the underlying patterns identified across all simulations (Fred, 2001). This method reduces the dependency on specific algorithmic parameters, thereby improving the robustness and reliability of the clustering results (Boongoen and Iam-On, 2018). Based on the computed co-association matrix, a new, more consensual partition is obtained in which elements assigned to the same clusters have a high probability of being together. Fig. 6.15 shows the computed co-association matrix.

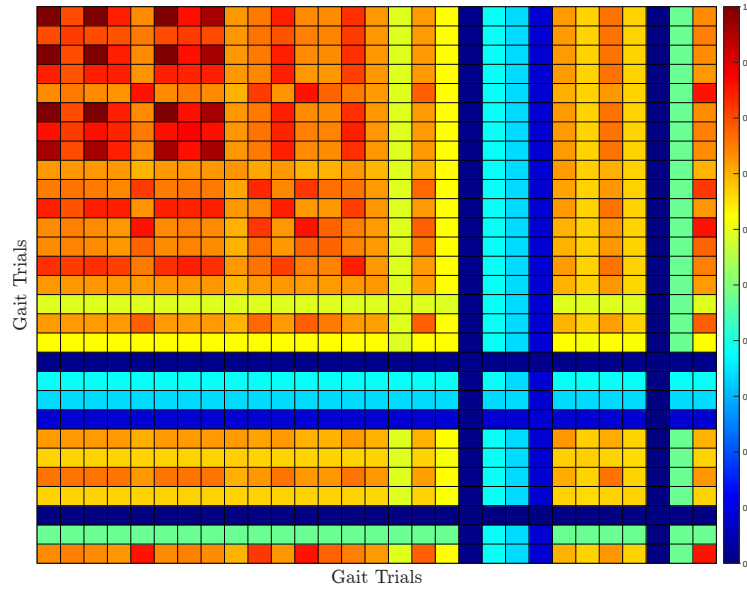


Fig. 6.15: Co-association matrix computed from multiple simulations with varying clustering algorithm parameters. Only partitions obtained using the Förstner distance-based dissimilarity matrix were considered.

When examining the co-association matrix, a clear division into regions similar to that presented in the Förstner distance-based dissimilarity matrix is noticed. The entries in the upper right region of the matrix have warmer colors, indicating that healthy subjects are more likely to belong to the same cluster. In contrast, the probability of healthy subjects and those with PD being assigned to the same cluster shows notable variations. Some pairs of elements show sufficient likelihood to be grouped together, while others clearly exhibit significant differences, resulting in a low probability of being assigned to the same cluster.

Fig. 6.16 shows a comparison between the actual distribution of healthy subjects and subjects with PD, along with the consensus clustering identified through the ensemble clustering approach.

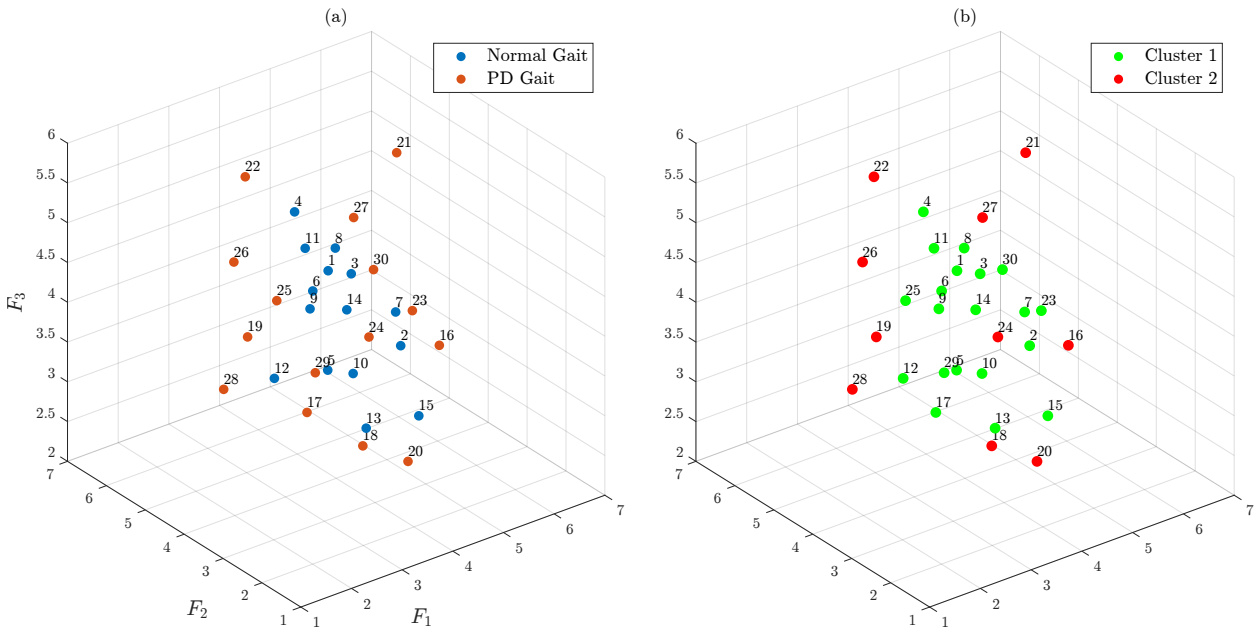


Fig. 6.16: Clusters identified from the Förstner distance-based network layout using the DBSCAN algorithm and clustering ensemble approach.

In analyzing the results of applying the clustering ensemble approach, we initially observed that the identified clusters closely align with the best clustering result obtained by exploring the NMI scores across the entire DBSCAN parameter grid. This result is promising, as it indicates the feasibility of achieving optimal clustering based solely on the exploration of the SC grid without relying on NMI scores.

After embedding gait-related data (time series) into a three-dimensional space and applying clustering algorithms, it is important to review the cases where the ICR-LLS methodology misclassifies subjects. In this study, subjects 17, 23, 25, 29, and 30 were incorrectly classified as healthy, despite being medically diagnosed with PD. Table 6.3 summarizes the demographic and medical information for all subjects diagnosed with PD, along with their corresponding diagnostic results from the ICR-LLS methodology. Misclassified subjects are highlighted in blue in the table.

Table 6.3: Information on subjects with medical diagnosed PD and the corresponding diagnosis resulting from using the ICR-LLS methodology.

Subject number	Gender	Age	Height (cm)	Weight (kg)	UPDRS-II	UPDRS-II (walking)	UPDRS-III	UPDRS-III (walking)	Hoehn & Yahr	Diagnosis
Subject 16	M	58.00	168	69	13	2	27	2	3	PD
Subject 17	F	53.00	170	62.55	2	1	10	0	1	Healthy
Subject 18	M	63.00	158.5	63.85	4	1	26	0	2	PD
Subject 19	M	68.00	169	68.9	10	3	48	2	3	PD
Subject 20	F	77.00	151.5	60.2	6	2	30	0	2	PD
Subject 21	M	46.00	171	61.5	10	2	6	1	2	PD
Subject 22	F	44.00	157	53.3	11	3	16	2	2	PD
Subject 23	M	57.00	173.5	57.7	5	1	24	1	2	Healthy
Subject 24	M	60.00	179	92.45	6	1	30	2	3	PD
Subject 25	M	81.00	154.5	65.65	7	1	28	2	3	Healthy
Subject 26	M	73.00	166	86	10	2	33	1	2	PD
Subject 27	F	61.00	158.5	60.5	13	3	61	2	3	PD
Subject 28	M	69.00	165	76.5	1	0	17	0	2	PD
Subject 29	M	74.00	179	94.55	4	1	29	1	3	Healthy
Subject 30	M	62.00	166.5	57.4	5	1	25	1	2	Healthy

Table 6.3 shows that when subjects are ranked from lowest to highest according to their UPDRS-II, UPDRS-II (walking), UPDRS-III, UPDRS-III (walking), and Hoehn & Yahr scores, none of the four subjects with the lowest scores correspond to the four misclassified subjects. This indicates that a more detailed analysis is needed. Notably, the scale most closely related to the results of the ICR-LLS methodology is the UPDRS-II (walking) scale. All subjects classified as healthy by the ICR-LLS methodology have a UPDRS-II (walking) score of 1, which is among the lowest scores, with the only lower score being 0, recorded by subject 28. Since the UPDRS-II scale assesses motor experiences of daily living, a score of 1 in walking suggests minimal disability, which might lead to misdiagnosis of PD due to subtle symptoms.

An examination of the Hoehn & Yahr stages reveals no clear correlation between these stages and the presence of subjects misclassified as healthy. Similarly, there is no strong correlation between the Hoehn & Yahr stages and the UPDRS-II scores. For example, subject 30 has a high Hoehn & Yahr stage but a low UPDRS-II score, indicating advanced PD with minimal observed motor impairment in gait. Likewise, UPDRS-III scores, which are derived from physician-performed motor examinations, also do not correlate with PD misdiagnosis. The scores of the misclassified subjects fall within the middle range, and none are in the top quartile (i.e., lowest scoring quartile). This disparity suggests that while UPDRS-II scores may contribute to misclassification, Hoehn & Yahr stages and UPDRS-III scores do not consistently affect the results of PD diagnosis when using the ICR-LLS methodology.

To evaluate the performance of the proposed methodology, a good practice, fundamental to enrich the research, is to compare it with similar studies found in the literature. However, to date, no study using the same approach in a comparable cohort of patients (i.e., Parkinson's patients and control subjects) has been identified.

Generally, the diagnosis of PD and other health disorders from gait analysis is based on classification techniques. This approach, based on ground truth knowledge, is popular for its simplicity and easily interpretable results. Novel studies employing this approach have focused on predicting disease severity, using advanced techniques such as deep learning classifiers to estimate scores such as UPDRS for PD (Aşuroğlu et al., 2018; Aşuroğlu and Oğul, 2022; Faiem et al., 2024). In contrast, this research employs a clustering-based approach, which distinguishes gait patterns based on the intrinsic characteristics of the data without relying on predefined labels. Because clustering methods are not well-suited for predicting disease severity, this aspect falls outside the scope of the present study. Instead, studies using cluster analysis of gait data for medical purposes have typically focused on identifying subgroups within homogeneous populations, such as control subjects, PD patients, or stroke patients (Kim et al., 2022).

Table 6.4: Comparison of the ICR-LLS methodology results with previous studies on PD classification

Authors	Acc. (%)	Sn. (%)	Sp. (%)	F-score	Gait parameters and features
Lee and Lim (2012)	74.3	81.60	73.80	no provided	GRFs
Daliri (2013)	91.20	91.71	89.92	no provided	GRFs
Ertuğrul et al. (2016)	88.80	88.90	82.20	no provided	GRFs
Açııcı et al. (2017)	98.00	99.10	95.70	0.98	GRFs
Aşuroğlu et al. (2018)	99.00	97.80	99.50	no provided	GRFs
Zeng et al. (2016)	98.80	98.92	98.63	no provided	GRFs
Zhao et al. (2018)	98.61	no provided	no provided	no provided	GRFs
Nancy Noella et al. (2019)	97.00	no provided	no provided	no provided	GRFs
Ghaderyan and Fathi (2021)	97.22	98.22	95.86	no provided	GRFs
Liu et al. (2021)	99.22	98.04	100	0.99	GRFs
Aşuroğlu and Oğul (2022)	99.5	98.70	99.10	0.99	GRFs
Ricciardi et al. (2019)	86.40	96.30	96.30	no provided	3D kinematics
Park et al. (2021)	98.10	no provided	no provided	no provided	3D kinematics
Ajay et al. (2018)	93.75	no provided	no provided	no provided	3D kinematics
Guayacán and Martínez (2021)	94.89	90.72	100	0.70	3D kinematics
Sun et al. (2024)	95.40	86.74	95.52	0.89	3D kinematics
Zhang et al. (2020)	82.70	83.80	82.10	no provided	3D kinematics
ICR-LLS methodology	83.33	66.67	100	0.80	ICR trajectories

A summary of the major PD diagnostic studies based on gait-related data is provided in Table 6.4. All of these studies employ classification methods for the detection of PD. To effectively compare our clustering-based approach with these existing methodologies, four performance metrics were calculated: accuracy, sensitivity, specificity, and F-score.

It is important to note that each study referenced in the table used different databases and that not all gait data in these studies are exclusively kinematic. In addition, the availability of preprocessed gait

kinematic data, which are essential for the ICR-LLS methodology, is limited. The database used in our study was recently published, and, to date, only three studies have used these data. These studies have focused primarily on estimating UPDRS scores or examining the effects of medication on gait patterns (Adeli et al., 2024; Los Angeles et al., 2024; Shida et al., 2023b). In contrast, many of the studies we compared are based on GRF data, which have been shown to produce superior results compared to kinematic data alone.

Although the performance metric scores of the ICR-LLS methodology are among the lowest compared to those of the selected studies, it is critical to recognize that the ICR-LLS methodology uses in-plane kinematic data rather than three-dimensional data. Incorporating additional in-plane ICR trajectories could potentially improve the results of the proposed methodology. However, a significant advantage of using only in-plane kinematic data is that it simplifies data acquisition by allowing the use of a single camera, as demonstrated in Chapter 5. This capability simplifies data collection and broadens the applicability of the methodology.

Finally, although the cluster-based nature of the ICR-LLS methodology complicates direct comparisons with existing studies, it highlights the innovative approach of this methodology. By embedding the gait-related time series in a three-dimensional space and focusing on characteristics derived from this operation, the ICR-LLS methodology offers a novel perspective for diagnosing potential driving problems that traditional methods may overlook.

Chapter 7

Conclusions and final remarks

This work aimed to investigate to what extent the ICR trajectories of the lower limb segment in the sagittal plane described during gait can be used as a source of diagnostic information to assess health problems that cause deviations from the typical gait pattern. The study includes a comprehensive review of gait-related data acquisition techniques and methodologies used to assess gait and diagnose disorders, as well as the study of the characteristic ICR trajectories of the thigh and tibia in the sagittal plane during gait. Finally, a novel methodology for the diagnosis of health problems affecting gait is proposed, based on the analysis of ICR trajectories, and used to discriminate parkinsonian gait patterns among healthy individuals.

Taking into account the tasks performed and the results obtained, it can be concluded that the established research objective has been achieved and the research hypothesis has been validated. Using the proposed methodology, we have effectively demonstrated that it is possible to differentiate individuals with abnormal gait from healthy individuals using only kinematic parameters, such as the ICR trajectory of the shank in the sagittal plane during walking.

The study carried out began with a detailed review of the parameters commonly measured and analyzed in gait studies, with a special emphasis on the kinematic parameters. The review was enriched with an overview of gait-related kinematic data acquisition techniques and a review of the methodologies commonly used in gait analysis to diagnose gait disorders. Among the review studies, several approaches for classification, clustering, and pattern analysis were investigated in order to propose an alternative methodology that produces good results using only kinematic gait-related data. From the review conducted, it is clear that current research on the diagnosis of health problems based on gait-related kinematic data has focused primarily on the study of angular movements of joints, leaving aside other kinematic parameters that can contain useful diagnostic information. Therefore, ICR motion, a notion commonly used in machine motion analysis, has been selected as the kinematic parameter analyzed in this study.

The ICR trajectories of the shank and thigh were studied in time, frequency, and phase space. To characterize the typical pattern of ICR trajectories, treadmill and overground gait trials were performed and data regarding the position in time of anatomical markers of the lower limbs were collected using an inertial MoCap system and a vision-based technique. The trajectories described by each anatomical marker during a single GC were recorded and segmented into the contained GCs, which were then averaged to obtain the typical trajectory of each marker during the GC. The typical trajectories of the markers were used to cal-

culate the ICR trajectories of the shank and thigh in the sagittal plane. In the temporal analysis of the ICR trajectories, the typical pattern of each trajectory along the GC and each gait phase was described by analyzing the data in Cartesian and polar coordinates. In the frequency analysis of the ICR trajectories, the energy distribution of the signals along discrete frequencies was obtained using FT and STFT. However, little relevant information was obtained from this analysis due to the presence of asymptotes, making them difficult to analyze. Finally, the trajectories were used to reconstruct the associated dynamical systems. Three systems consisting of shank, thigh, and shank-thigh were considered, respectively. To effectively visualize and characterize the reconstructed state spaces, the RPs were calculated, providing a detailed characterization of each individual system.

As a result of the study, it was found that strict use of kinematic parameters in gait analysis is sufficient to extract relevant information on the health status of individuals, at least in the situations studied. Specifically, the use of the ICR trajectories of the shank proved to be a reliable parameter to characterize gait, despite the complexities involved in its analysis because of the non-smooth nature of the ICR trajectories and the presence of asymptotes.

To diagnose health problems that affect gait, a five-stage methodology has been proposed. In the first stage, gait-related kinematic data consisting of the trajectory of anatomical markers of the subjects' lower extremity are collected. To acquire the data, treadmill and overground gait trials are performed during which the movement of the subjects tested is recorded using the MoCap system or video-based techniques. This stage is followed by the calculation of ICR trajectories from the collected data and the assembly of a multivariate time series defining the subject's gait. Each multivariate time series is made up of the ICR trajectories of the left and right legs on the vertical axis and on the horizontal axis. In the third stage, the multivariate time series are compared pairwise using a selected dissimilarity measure. As a result of the pairwise comparisons performed, the dissimilarity matrix is obtained, which contains information on the degree of similarity between the different gait trials performed in the first stage, and therefore of the gait itself, of the subjects tested. In the fourth stage, using the information contained in the dissimilarity matrix, the multivariate time series, representing the gait of each subject, are embedded in a three-dimensional space. Each multivariate time series is transformed into a node whose spatial location with respect to the other nodes is determined by the information contained in the dissimilarity matrix, i.e., the distance between two nodes of the formed network is proportional to the dissimilarity between the gait of two subjects. The task of embedding the time series in a three-dimensional space is carried out using a self-organizing method based on force-directed graph drawing. This method involves representing the arrangement of nodes as a mechanical system in which negligible masses (nodes) are linked by springs whose stiffness constant is proportional to the dissimilarity between the connected nodes. Finally, in the last stage, which is optional, a clustering algorithm is used to identify the clusters formed in the three-dimensional space in which the gait-related data were embedded. The selected algorithm was DBSCAN, due to its simple principle of operation and implementation and the fact that it handles noise within the data set relatively well.

The sensitivity of the proposed methodology was evaluated considering a single subject who underwent treadmill and overground gait trials with different levels of added difficulty. A total of four gait trials were performed. During gait trials, a heel was added to the subject's shoe and its height was increased with each

trial. The results showed that there is a clear relationship between the increase in gait difficulty and a scalar parameter, called the global gait characteristic, derived from the results applying the ICR-LLS methodology. An additional analysis was performed in which gait-related data from several subjects with different levels of added difficulty were recorded and processed using ICR-LLS to discriminate each level of difficulty contained in the data set. Contrary to expectations, ICR-LLS appeared to differentiate subjects rather than differentiating the levels of gait difficulty, which is a promising result if what is desired is a methodology for person recognition using gait patterns.

Finally, the ICR-LLS methodology was employed to discriminate gait patterns in a more realistic context. This involved using a data set that contains gait-related data from PD patients along with an age-matched data set of healthy subjects. The methodology yielded promising results, achieving a maximum NMI score of 0.479 when using the Förstner distance. Although the results were not perfect, they are encouraging, as no healthy subjects were misclassified and misclassified subjects with PD had relatively lower PD scores, particularly in UPDRS-II. This outcome suggests that the proposed methodology is effective in differentiating normal gait from pathological gait. However, more studies are necessary to achieve perfect discrimination performance, particularly if it is aimed at addressing more challenging tasks, such as distinguishing between different types of gait abnormalities.

Within the framework of the study carried out and presented in this dissertation, the author's main achievements can be summarized as follows:

- ✓ Review of current state-of-the-art techniques used for the acquisition of gait-related kinematic data, and the approaches used to diagnose gait disorders in clinical settings and in specialized gait laboratories.
- ✓ Implementation of algorithms to segment GCs contained in an acquired gait-related time series suitable for segmenting GCs from treadmill and overground gait-related time series, and algorithms to determine the position of the ICR of body segments based on data collected from inertial-based and vision-based data acquisition techniques.
- ✓ Creation of a data set containing kinematic data on the movement of body landmarks and the ICR trajectories of the shank and thigh in the sagittal plane during treadmill and overground gait trials.
- ✓ Characterization and comparison of ICR trajectories of the thigh and shank, calculated from data collected from treadmill and overground gait trials and using two different data acquisition techniques, in the domain of time, frequency, and phase space.
- ✓ Dynamic study of gait by analyzing the phase spaces of three dynamic systems, derived from (1) ICR trajectories of the thigh, (2) ICR trajectories of the shank, and (3) ICR trajectories of the thigh and shank, respectively, using RP and RQA.
- ✓ Proposition of a novel methodology for the diagnosis of gait-related health problems, based on pairwise comparisons between multivariate time series and embedding of the data in a three-dimensional space for subsequent clustering.

- ✓ Performance evaluation of the proposed methodology using nine different dissimilarity measures and different levels of gait difficulties intentionally induced to individuals during gait trials to simulate gait abnormalities.
- ✓ An application of the proposed methodology to discriminate Parkinsonian gait from normal gait using a public data set.

The application of the ICR-LLS methodology offers several advantages in the diagnosis of health conditions that affect gait. It efficiently handles time-dependent gait-related data, producing easily interpretable visual results crucial to clinical decision-making processes. Furthermore, it relies strictly on gait-related kinematic data, eliminating the need for additional equipment for the acquisition of kinetic or EMG data acquisition, which distinguishes it from traditional gait analysis performed in specialized laboratories. Consequently, the ICR-LLS methodology is presented as a quantitative approach, an alternative to practices in clinical settings that rely heavily on the expertise of clinicians.

To enrich the research conducted and advance the field of gait analysis, several promising avenues for future research can be explored. A primary focus should be to refine the ICR-LLS methodology to improve its ability to differentiate between various gait abnormalities beyond Parkinsonian gait. This refinement could be achieved by integrating additional gait features and combining ICR data with kinetic or EMG measurements. This comprehensive approach would allow for a more precise analysis, improving diagnostic precision, and providing more insight into the mechanisms underlying different gait abnormalities. It is also crucial to extend the ICR-LLS methodology to practical, real-world settings. Evaluation of its performance in clinical settings and rehabilitation centers will provide valuable information on its robustness and utility in the diagnosis and follow-up of a variety of gait disorders. This real-world evaluation will help determine the efficacy of the methodology beyond controlled experimental conditions. Finally, it is essential to broaden the scope of the research to include a more diverse population sample. By including participants of different ages and types of gait disorders, researchers can assess the generalizability of the ICR-LLS methodology and refine it to ensure its efficacy in a wider range of individuals. The study of ICR trajectories throughout the gait cycle opens new possibilities for new methodologies not only for the diagnosis of health problems that affect gait, but also for tasks such as the recognition of individuals by their gait.

Similar research carried out by the author, focusing on biomechanical analysis of the human body for the diagnosis of gait disorders, the development of assistive devices, and energy harvesting has been published in articles ([Molina Arias et al., 2021a,b](#); [Molina and Iwaniec, 2022](#); [Molina Arias et al., 2023b,a](#)) and presented at numerous international conferences.

Bibliography

- Abbass, S. J. and Abdulrahman, G. (2014). Kinematic analysis of human gait cycle. *College of Engineering Journal (NUCEJ)* 16, 208–222
- Açııcı, K., Erdaş, Ç. B., Aşuroğlu, T., Toprak, M. K., Erdem, H., and Oğul, H. (2017). A Random Forest Method to Detect Parkinson’s Disease via Gait Analysis. In *Engineering Applications of Neural Networks*, eds. G. Boracchi, L. Iliadis, C. Jayne, and A. Likas (Cham: Springer International Publishing), vol. 744 of *Communications in Computer and Information Science*. 609–619. doi:10.1007/978-3-319-65172-9_51
- Adams, J. M. and Cerny, K. (2018). *Observational Gait Analysis : A Visual Guide* (SLACK Incorporated), 1st edn.
- Adeli, V., Mehraban, S., Ballester, I., Zarghami, Y., Sabo, A., Iaboni, A., et al. (2024). Benchmarking Skeleton-based Motion Encoder Models for Clinical Applications: Estimating Parkinson’s Disease Severity in Walking Sequences. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)* (IEEE), 1–10. doi:10.1109/FG59268.2024.10581933
- Aiyangar, A., Zheng, L., Anderst, W., and Zhang, X. (2017). Instantaneous centers of rotation for lumbar segmental extension in vivo. *Journal of Biomechanics* 52, 113–121. doi:10.1016/j.jbiomech.2016.12.021
- Ajay, J., Song, C., Wang, A., Langan, J., Li, Z., and Xu, W. (2018). A pervasive and sensor-free Deep Learning system for Parkinsonian gait analysis. In *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)* (IEEE), March, 108–111. doi:10.1109/BHI.2018.8333381
- Alfayeed, S. M. and Saini, B. S. (2021). Human Gait Analysis Using Machine Learning: A Review. In *Proceedings of the 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*. 550–554. doi:10.1109/ICCIKE51210.2021.9410678
- Alharthi, A. S., Yunas, S. U., and Ozanyan, K. B. (2019). Deep learning for monitoring of human gait: A review. *IEEE Sensors Journal* 19, 9575–9591. doi:10.1109/JSEN.2019.2928777
- Allen, N. E., Schwarzel, A. K., and Canning, C. G. (2013). Recurrent Falls in Parkinson’s Disease: A Systematic Review. *Parkinson’s Disease* 2013, 1–16. doi:10.1155/2013/906274
- Alqahtani, A., Ali, M., Xie, X., and Jones, M. W. (2021). Deep Time-Series Clustering: A Review. *Electronics* 10, 3001. doi:10.3390/electronics10233001

- Amboni, M., Barone, P., and Hausdorff, J. M. (2013). Cognitive contributions to gait and falls: Evidence and implications. *Movement Disorders* 28, 1520–1533. doi:10.1002/mds.25674
- Amelio, A. and Pizzuti, C. (2015). Is Normalized Mutual Information a Fair Measure for Comparing Community Detection Methods? In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. 1584–1585. doi:10.1145/2808797.2809344
- Aritan, S. (2015). Biomechanical Measurement Methods to Analyze the Mechanisms of Sport Injuries. In *Sports Injuries: Prevention, Diagnosis, Treatment and Rehabilitation* (Springer Berlin, Heidelberg). 2nd edn., 3085–3096. doi:10.1007/978-3-642-36569-0
- Aşuroğlu, T., Açıcı, K., Berke Erdaş, Ç., Kılınç Toprak, M., Erdem, H., and Oğul, H. (2018). Parkinson's disease monitoring from gait analysis via foot-worn sensors. *Biocybernetics and Biomedical Engineering* 38, 760–772. doi:10.1016/j.bbe.2018.06.002
- Aşuroğlu, T. and Oğul, H. (2022). A deep learning approach for parkinson's disease severity assessment. *Health and Technology* 12, 943–953. doi:10.1007/s12553-022-00698-z
- Backlund, A. (2000). The definition of system. *Kybernetes* 29, 444–451. doi:10.1108/03684920010322055
- Baker, R. (2006). Gait analysis methods in rehabilitation. *Journal of NeuroEngineering and Rehabilitation* 3, 1–10. doi:10.1186/1743-0003-3-4
- Barre, A. and Armand, S. (2014). Biomechanical ToolKit: Open-source framework to visualize and process biomechanical data. *Computer Methods and Programs in Biomedicine* 114, 80–87. doi:10.1016/j.cmpb.2014.01.012
- Bayat, A., Bayat, A. H., and Ghasemi, A. S. (2017). Classifying Human Walking Patterns using Accelerometer Data from Smartphone. *International Journal of Computer Science and Mobile Computing* 612, 78–83
- Bonnefoy-Mazure, A. and Armand, S. (2015). Normal Gait. In *Orthopedic Management of Children with Cerebral Palsy A Comprehensive Approach* (Nova Science Publisher Inc). 645–657
- Boongoen, T. and Iam-On, N. (2018). Cluster ensembles: A survey of approaches with recent extensions and applications. *Computer Science Review* 28, 1–25. doi:10.1016/j.cosrev.2018.01.003
- Boulgouris, N., Plataniotis, K., and Hatzinakos, D. (2004). Gait recognition using dynamic time warping. In *Proceedings of the IEEE 6th Workshop on Multimedia Signal Processing (MMSP)*. January, 263–266. doi:10.1109/MMSP.2004.1436543
- Bradley, E. (1999). Intelligent Data Analysis. In *Analysis of Time Series* (Springer Berlin Heidelberg). 146–167. doi:10.1007/978-3-662-03969-4_5
- BTS Bioengineering (n.d.). BTS GaitLab Setup. <https://www.btsbioengineering.com/products/bts-gaitlab/>. [Accessed 29 January 2024]

- Cassisi, C., Montalto, P., Aliotta, M., Cannata, A., and Pulvirenti, A. (2012). Similarity Measures and Dimensionality Reduction Techniques for Time Series Data Mining. In *Advances in Data Mining Knowledge Discovery and Applications* (InTech). 71–96. doi:10.5772/49941
- Cavazza, J., Zunino, A., Biagio, M. S., and Murino, V. (2016). Kernelized covariance for action recognition. In *Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR)*. 408–413. doi:10.1109/ICPR.2016.7899668
- Chantraine, F., Schreiber, C., Pereira, J. A. C., Kaps, J., and Dierick, F. (2022). Classification of Stiff-Knee Gait Kinematic Severity after Stroke Using Retrospective k-Means Clustering Algorithm. *Journal of Clinical Medicine* 11, 6270. doi:10.3390/jcm11216270
- Chen, M., Huang, B., and Xu, Y. (2007). Human Abnormal Gait Modeling via Hidden Markov Model. In *Proceedings of the 2007 International Conference on Information Acquisition (ICIA)*. 517–522. doi:10.1109/ICIA.2007.4295787
- Chen, S., Lach, J., Lo, B., and Yang, G. Z. (2016). Toward Pervasive Gait Analysis With Wearable Sensors: A Systematic Review. *IEEE Journal of Biomedical and Health Informatics* 20, 1521–1537. doi:10.1109/JBHI.2016.2608720
- Cicirelli, G., Impedovo, D., Dentamaro, V., Marani, R., Pirlo, G., and D’Orazio, T. R. (2022). Human Gait Analysis in Neurodegenerative Diseases: A Review. *IEEE Journal of Biomedical and Health Informatics* 26, 229–242. doi:10.1109/JBHI.2021.3092875
- Daliri, M. R. (2013). Chi-square distance kernel of the gaits for the diagnosis of Parkinson’s disease. *Biomedical Signal Processing and Control* 8, 66–70. doi:10.1016/j.bspc.2012.04.007
- Daszykowski, M., Walczak, B., Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. 565–580. doi:10.1016/B978-0-444-64165-6.03005-6
- DeCann, B., Ross, A., and Culp, M. (2014). On Clustering Human Gait Patterns. In *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR)*. 1794–1799. doi:10.1109/ICPR.2014.315
- Delisa, J. A. (1998). *Gait analysis in the science of rehabilitation* (Department of Veterans Affairs)
- Diab, D. M., AsSadhan, B., Binsalleeh, H., Lambbotharan, S., Kyriakopoulos, K. G., and Ghafir, I. (2019). Anomaly Detection Using Dynamic Time Warping. In *Proceedings of the 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*. 193–198. doi:10.1109/CSE/EUC.2019.00045
- Disease, M. D. S. T. F. o. R. S. f. P. (2003). The Unified Parkinson’s Disease Rating Scale (UPDRS): Status and recommendations. *Movement Disorders* 18, 738–750. doi:10.1002/mds.10473

- Ebersbach, G., Sojer, M., Miller, J., Heijmenberg, M., and Poewe, W. (2000). Sociocultural differences in gait. *Movement Disorders* 15, 1145–1147. doi:10.1002/1531-8257(200011)15:6<1145::AID-MDS1013>3.0.CO;2-C
- Ergüner Özkoç, E. (2021). Clustering of Time-Series Data. In *Data Mining - Methods, Applications and Systems* (IntechOpen). 13. doi:10.5772/intechopen.84490
- Ertuğrul, Ö. F., Kaya, Y., Tekin, R., and Almalı, M. N. (2016). Detection of Parkinson's disease by Shifted One Dimensional Local Binary Patterns from gait. *Expert Systems with Applications* 56, 156–163. doi:10.1016/j.eswa.2016.03.018
- Faiem, N., Asuroglu, T., Acici, K., Kallonen, A., and van Gils, M. (2024). Assessment of Parkinson's Disease Severity Using Gait Data: A Deep Learning-Based Multimodal Approach. In *Digital Health and Wireless Solutions*, eds. M. Särestöniemi, P. Keikhosrokiani, D. Singh, E. Harjula, A. Tiulpin, M. Jansson, M. Isomursu, M. van Gils, S. Saarakkala, and J. Reponen (Cham: Springer Nature Switzerland), vol. 2084 of *Communications in Computer and Information Science*. 29–48. doi:10.1007/978-3-031-59091-7_3
- Faloutsos, C., Ranganathan, M., and Manolopoulos, Y. (1994). Fast subsequence matching in time-series databases. In *Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data*. 419–429. doi:10.1145/191839.191925
- Fan, C., Chen, M., Wang, X., Wang, J., and Huang, B. (2021). A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data. *Frontiers in Energy Research* 9, 1–17. doi:10.3389/fenrg.2021.652801
- Ferber, R., Osis, S. T., Hicks, J. L., and Delp, S. L. (2016). Gait biomechanics in the era of data science. *Journal of Biomechanics* 49, 3759–3761. doi:10.1016/j.jbiomech.2016.10.033
- Frankel, V. H., Burstein, A. H., and Brooks, D. B. (1971). Biomechanics of Internal Derangement of the Knee. *The Journal of Bone & Joint Surgery* 53, 87–129. doi:10.2106/00004623-197153050-00010
- Fred, A. (2001). Finding Consistent Clusters in Data Partitions. In *Lecture Notes in Computer Science* (Springer Berlin Heidelberg), vol. 2096. 309–318. doi:10.1007/3-540-48219-9_31
- Fu, T. C. (2011). A review on time series data mining. *Engineering Applications of Artificial Intelligence* 24, 164–181. doi:10.1016/j.engappai.2010.09.007
- Förstner, W. and Moonen, B. (2003). A metric for covariance matrices. In *Geodesy-The Challenge of the 3rd Millennium* (Springer Berlin Heidelberg). 299–309. doi:10.1007/978-3-662-05296-9_31
- Gao, C. X., Dwyer, D., Zhu, Y., Smith, C. L., Du, L., Folia, K. M., et al. (2023). An overview of clustering methods with guidelines for application in mental health research. *Psychiatry Research* 327, 115265. doi:10.1016/j.psychres.2023.115265
- Ghaderyan, P. and Fathi, G. (2021). Inter-limb time-varying singular value : A new gait feature for Parkinson's disease detection and stage classification. *Measurement* 177, 109249. doi:10.1016/j.measurement.2021.109249

- Giladi, N. (2019). Gait Disorders: Clinical Overview. <https://www.movementdisorders.org/MDS/About/Movement-Disorder-Overviews/Gait-Disorders.htm>. [Accessed 29 January 2024]
- Guayacán, L. C. and Martínez, F. (2021). Visualising and quantifying relevant parkinsonian gait patterns using 3D convolutional network. *Journal of Biomedical Informatics* 123, 103935. doi:10.1016/j.jbi.2021.103935
- Guo, Y., Yang, J., Liu, Y., Chen, X., and Yang, G. Z. (2022). Detection and assessment of Parkinson's disease based on gait analysis: A survey. *Frontiers in Aging Neuroscience* 14. doi:10.3389/fnagi.2022.916971
- Hamilton, J. D. (1994). The State-Space Representation of a Dynamic System. In *Time Series Analysis* (Princeton University Press). 372–377. doi:10.1515/9780691218632-111
- Harris, G. F. and Wertsch, J. J. (1994). Procedures for gait analysis. *Archives of Physical Medicine and Rehabilitation* 75, 216–225. doi:10.1016/0003-9993(94)90399-9
- Healy, A., Farmer, S., Pandyan, A., and Chockalingam, N. (2018). A systematic review of randomised controlled trials assessing effectiveness of prosthetic and orthotic interventions. *PLoS ONE* 13, e0192094. doi:10.1371/journal.pone.0192094
- Hoehn, M. M. and Yahr, M. D. (1998). Parkinsonism onset, progression, and mortality. *Neurology* 50, 318–318. doi:10.1212/WNL.50.2.318
- Holden, M. K., Gill, K. M., and Magliozzi, M. R. (1986). Gait assessment for neurologically impaired patients. Standards for outcome assessment. *Physical therapy* 66, 1530–1539. doi:10.1093/ptj/66.10.1530
- Howcroft, J., Kofman, J., and Lemaire, E. D. (2017). Feature selection for elderly faller classification based on wearable sensors. *Journal of NeuroEngineering and Rehabilitation* 14, 1–12. doi:10.1186/s12984-017-0255-9
- Hu, Y. (2006). Efficient and High Quality Force-Directed Graph. *The Mathematica Journal* 10, 37–71
- Hulleck, A. A., Menoth Mohan, D., Abdallah, N., El Rich, M., and Khalaf, K. (2022). Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies. *Frontiers in Medical Technology* 4. doi:10.3389/fmedt.2022.901331
- Kantz, H. and Schreiber, T. (2003). *Nonlinear Time Series Analysis* (Cambridge University Press), 2nd edn. doi:10.1017/CBO9780511755798
- Kanwal, S. and Asghar, S. (2021). Speech Emotion Recognition Using Clustering Based GA-Optimized Feature Set. *IEEE Access* 9, 125830–125842. doi:10.1109/ACCESS.2021.3111659
- Kennel, M. B., Brown, R., and Abarbanel, H. D. I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A* 45, 3403–3411. doi:10.1103/PhysRevA.45.3403

- Khera, P. and Kumar, N. (2020). Role of machine learning in gait analysis: a review. *Journal of Medical Engineering and Technology* 44, 441–467. doi:10.1080/03091902.2020.1822940
- Khokhlova, M., Migniot, C., and Dipanda, A. (2018). Kinematic Covariance Based Abnormal Gait Detection. In *Proceedings of the 14th International Conference on Signal Image Technology and Internet Based Systems (SITIS)*. 691–696. doi:10.1109/SITIS.2018.00111
- Khokhlova, M., Migniot, C., Morozov, A., Sushkova, O., and Dipanda, A. (2019). Normal and pathological gait classification LSTM model. *Artificial Intelligence in Medicine* 94, 54–66. doi:10.1016/j.artmed.2018.12.007
- Kidziński, Ł., Yang, B., Hicks, J. L., Rajagopal, A., Delp, S. L., and Schwartz, M. H. (2020). Deep neural networks enable quantitative movement analysis using single-camera videos. *Nature Communications* 11, 1–10. doi:10.1038/s41467-020-17807-z
- Kim, H., Kim, Y.-H., Kim, S.-J., and Choi, M.-T. (2021). Gait Clustering Analysis in Patients after Stroke using Gait Kinematics Data. In *Proceedings of the 2021 21st International Conference on Control, Automation and Systems (ICCAS)*. 1405–1409. doi:10.23919/ICCAS52745.2021.9649908
- Kim, H., Kim, Y. H., Kim, S. J., and Choi, M. T. (2022). Pathological gait clustering in post-stroke patients using motion capture data. *Gait and Posture* 94, 210–216. doi:10.1016/j.gaitpost.2022.03.007
- Kim, S. M., Kim, D. H., Yang, Y., Ha, S. W., and Han, J. H. (2018). Gait Patterns in Parkinson's Disease with or without Cognitive Impairment. *Dementia and Neurocognitive Disorders* 17, 57. doi:10.12779/dnd.2018.17.2.57
- Kim, W. and Kim, Y. (2021). Abnormal gait recognition based on integrated gait features in machine learning. In *Proceedings of the 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*. 1683–1688. doi:10.1109/COMPSAC51774.2021.00251
- Kobourov, S. G. (2012). Spring Embedders and Force Directed Graph Drawing Algorithms. [Preprint]
- Krishnan, S. (2021). Advanced analysis of biomedical signals. In *Biomedical Signal Analysis for Connected Healthcare* (Elsevier). 157–222. doi:10.1016/B978-0-12-813086-5.00003-7
- Kumar, M. S. N. and Babu, R. V. (2012). Human gait recognition using depth camera. In *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*. 1–6. doi:10.1145/2425333.2425353
- Lee, S. and Shin, S. (2014). Gait Signal Analysis with Similarity Measure. *The Scientific World Journal* 2014, 1–8. doi:10.1155/2014/136018
- Lee, S.-h. and Lim, J. S. (2012). Parkinson's disease classification using gait characteristics and wavelet-based feature extraction. *Expert Systems with Applications* 39, 7338–7344. doi:10.1016/j.eswa.2012.01.084

- Li, Q., Wang, Y., Sharf, A., Cao, Y., Tu, C., Chen, B., et al. (2018). Classification of gait anomalies from kinect. *Visual Computer* 34, 229–241. doi:10.1007/s00371-016-1330-0
- Liu, X., Li, W., Liu, Z., Du, F., and Zou, Q. (2021). A dual-branch model for diagnosis of Parkinson's disease based on the independent and joint features of the left and right gait. *Applied Intelligence* 51, 7221–7232. doi:10.1007/s10489-020-02182-5
- Los Angeles, E., de Oliveira, C. E. N., Cupertino, L., Shokur, S., Bouri, M., and Coelho, D. B. (2024). Effect of disease, freezing of gait, and dopaminergic medication in the biomechanics of trunk and upper limbs in the gait of Parkinson's disease. *Human Movement Science* 96, 103242. doi:10.1016/j.humov.2024.103242
- Lyons, R. (2010). *Understanding Digital Signal Processing* (Prentice Hall), 3rd edn.
- Mahlknecht, P., Kiechl, S., Bloem, B. R., Willeit, J., Scherfler, C., Gasperi, A., et al. (2013). Prevalence and Burden of Gait Disorders in Elderly Men and Women Aged 60-97 Years: A Population-Based Study. *PLoS ONE* 8, 1–7. doi:10.1371/journal.pone.0069627
- Mannini, A., Trojaniello, D., Cereatti, A., and Sabatini, A. M. (2016). A machine learning framework for gait classification using inertial sensors: Application to elderly, post-stroke and huntington's disease patients. *Sensors (Switzerland)* 16. doi:10.3390/s16010134
- Marsden, C. D. (1994). Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry* 57, 672–681. doi:10.1136/jnnp.57.6.672
- Marwan, N., Carmen Romano, M., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports* 438, 237–329. doi:10.1016/j.physrep.2006.11.001
- Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., and Kurths, J. (2002). Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical Review E* 66, 026702. doi:10.1103/PhysRevE.66.026702
- Mehrizi, R., Peng, X., Zhang, S., Liao, R., and Li, K. (2019). Automatic Health Problem Detection from Gait Videos Using Deep Neural Networks. [Preprint]
- Meyers, A., Buqammaz, M., and Yang, H. (2020). Cross-recurrence analysis for pattern matching of multi-dimensional physiological signals. *Chaos* 30, 123125. doi:10.1063/5.0030838
- Mohamed, A. A., Baba, J., Beyea, J., Landry, J., Sexton, A., and McGibbon, C. A. (2012). Comparison of strain-gage and fiber-optic goniometry for measuring knee kinematics during activities of daily living and exercise. *Journal of Biomechanical Engineering* 134, 084502. doi:10.1115/1.4007094
- Molina, L. and Iwaniec, M. (2022). Lower limb models used for biomechanical analysis of human walking. *MATEC Web of Conferences* 357, 03006. doi:10.1051/mateconf/202235703006

- Molina Arias, L., Iwaniec, J., and Iwaniec, M. (2021a). Modeling and Analysis of the Power Conditioning Circuit for an Electromagnetic Human Walking-Induced Energy Harvester. *Energies* 14, 3367. doi: 10.3390/en14123367
- Molina Arias, L., Iwaniec, M., and Iwaniec, J. (2023a). Investigation Into the Effect of Joint Clearance on the Dynamics of a Biomechanical Energy Harvesting System. *IEEE Access* 11, 24077–24086. doi: 10.1109/ACCESS.2023.3254207
- Molina Arias, L., Iwaniec, M., Pirowska, P., Smoleń, M., and Augustyniak, P. (2023b). Head and Voice-Controlled Human-Machine Interface System for Transhumeral Prosthesis. *Electronics* 12, 4770. doi: 10.3390/electronics12234770
- Molina Arias, L., Narina, R., and Iwaniec, M. (2021b). Modeling of Linear Electromagnetic Damper for Walking Assistance and Energy Harvesting. In *2021 IEEE XVIIth International Conference on the Perspective Technologies and Methods in MEMS Design (MEMSTECH)* (IEEE), 111–115. doi: 10.1109/MEMSTECH53091.2021.9468001
- Montgomery, S. C., Moorehead, J. D., Davidson, J. S., Lowe, D., and Dangerfield, P. H. (1998). A new technique for measuring the rotational axis pathway of a moving knee. *Knee* 5, 289–295. doi:10.1016/S0968-0160(97)10029-1
- Movella Inc. (n.d.). MVN AWINDA. <https://www.movella.com/products/motion-capture/xsens-mvn-awinda>. [Accessed 29 January 2024]
- Mulroy, S., Gronley, J., Weiss, W., Newsam, C., and Perry, J. (2003). Use of cluster analysis for gait pattern classification of patients in the early and late recovery phases following stroke. *Gait & Posture* 18, 114–125. doi:10.1016/S0966-6362(02)00165-0
- Muro-de-la Herran, A., García-Zapirain, B., and Méndez-Zorrilla, A. (2014). Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors (Switzerland)* 14, 3362–3394. doi:10.3390/s140203362
- Mutch, W. J., Dingwall-Fordyce, I., Downie, A. W., Paterson, J. G., and Roy, S. K. (1986). Parkinson's disease in a Scottish city. *British Medical Journal* 292, 534–536. doi:10.1136/bmj.292.6519.534
- Nancy Noella, R. S., Gupta, D., and Priyadarshini, J. (2019). Diagnosis of Parkinson's disease using Gait Dynamics and Images. *Procedia Computer Science* 165, 428–434. doi:10.1016/j.procs.2020.01.002
- Ng, H., Tong, H.-L., Tan, W.-H., and Abdullah, J. (2011). Gait Classification by Support Vector Machine. In *Communications in Computer and Information Science* (Springer Berlin, Heidelberg), vol. 179. 623–636. doi:10.1007/978-3-642-22170-5_54
- Nguyen, N. and Caruana, R. (2007). Consensus Clusterings. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)* (IEEE), 607–612. doi:10.1109/ICDM.2007.73

- Nithyakani, P. and Ferni Ukrit, M. (2020). The systematic review on gait analysis: Trends and developments. *European Journal of Molecular & Clinical Medicine* 7, 1636–1654
- Nuckols, R. W., Takahashi, K. Z., Farris, D. J., Mizrachi, S., Riemer, R., and Sawicki, G. S. (2020). Mechanics of walking and running up and downhill: A joint-level perspective to guide design of lower-limb exoskeletons. *PLoS ONE* 15, 1–21. doi:10.1371/journal.pone.0231996
- Park, H., Shin, S., Youm, C., Cheon, S.-M., Lee, M., and Noh, B. (2021). Classification of Parkinson's disease with freezing of gait based on 360° turning analysis using 36 kinematic features. *Journal of NeuroEngineering and Rehabilitation* 18, 177. doi:10.1186/s12984-021-00975-4
- Pathokinesiology Laboratory and Department of Physical Therapy (2001). *Observational gait analysis* (Rancho Los Amigos National Rehabilitation Center), 4th edn.
- Perry, J. (1992). *Gait Analysis: Normal and Pathological Function*, vol. 12 (SLACK Incorporated)
- Phinyomark, A., Petri, G., Ibáñez-Marcelo, E., Osis, S. T., and Ferber, R. (2018). Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions. *Journal of Medical and Biological Engineering* 38, 244–260. doi:10.1007/s40846-017-0297-2
- Pirker, W. and Katzenschlager, R. (2017). Gait disorders in adults and the elderly: A clinical guide. *Wiener Klinische Wochenschrift* 129, 81–95. doi:10.1007/s00508-016-1096-4
- Prakash, C., Mittal, A., Kumar, R., and Mittal, N. (2015). Identification of spatio-temporal and kinematics parameters for 2-D optical gait analysis system using passive markers. In *Proceedings of the 2015 International Conference on Advances in Computer Engineering and Applications (ICACEA)*. 143–149. doi:10.1109/ICACEA.2015.7164683
- Ramakrishnan, T., Lahiff, C. A., and Reed, K. B. (2018). Comparing gait with multiple physical asymmetries using consolidated metrics. *Frontiers in Neurorobotics* 12, 1–12. doi:10.3389/fnbot.2018.00002
- Rathinam, C., Bateman, A., Peirson, J., and Skinner, J. (2014). Observational gait assessment tools in paediatrics - A systematic review. *Gait & posture* 40. doi:10.1016/j.gaitpost.2014.04.187
- Reuleaux, F. (1876). *The Kinematics of Machinery: Outlines of a Theory of Machines* (Dover Publications)
- Ricciardi, C., Amboni, M., De Santis, C., Improta, G., Volpe, G., Iuppariello, L., et al. (2019). Using gait analysis' parameters to classify Parkinsonism: A data mining approach. *Computer Methods and Programs in Biomedicine* 180, 105033. doi:10.1016/j.cmpb.2019.105033
- Rodríguez-Licea, M. A., Perez-Pinal, F. J., Nuñez-Pérez, J. C., and Sandoval-Ibarra, Y. (2019). On the n-dimensional phase portraits. *Applied Sciences (Switzerland)* 9, 1–19. doi:10.3390/app9050872
- Rubenstein, L. Z., Solomon, D. H., Roth, C. P., Young, R. T., Shekelle, P. G., Chang, J. T., et al. (2004). Detection and management of falls and instability in vulnerable elders by community physicians. *Journal of the American Geriatrics Society* 52, 1527–1531. doi:10.1111/j.1532-5415.2004.52417.x

- Rupprechter, S., Morinan, G., Peng, Y., Foltynie, T., Sibley, K., Weil, R. S., et al. (2021). A Clinically Interpretable Computer-Vision Based Method for Quantifying Gait in Parkinson's Disease. *Sensors* 21, 5437. doi:10.3390/s21165437
- Salarian, A., Russmann, H., Vingerhoets, F. J. G., Dehollain, C., Blanc, Y., Burkhard, P. R., et al. (2004). Gait assessment in Parkinson's disease: toward an ambulatory system for long-term monitoring. *IEEE transactions on bio-medical engineering* 51, 1434–1443. doi:10.1109/TBME.2004.827933
- Schafer, R. (2011). What Is a Savitzky-Golay Filter? [Lecture Notes]. *IEEE Signal Processing Magazine* 28, 111–117. doi:10.1109/MSP.2011.941097
- Schepers, M., Giuberti, M., and Bellusci, G. (2018). *Xsens MVN: Consistent Tracking of Human Motion Using Inertial Sensing*. Tech. Rep. March, Xsens Technologies. doi:10.13140/RG.2.2.22099.07205
- Shida, T. K. F., Costa, T. M., de Oliveira, C. E. N., de Castro Treza, R., Hondo, S. M., Angeles, E. L., et al. (2023a). A public data set of walking full-body kinematics and kinetics in individuals with parkinson's disease. *Frontiers in Neuroscience* 17, 1–8. doi:10.3389/fnins.2023.992585
- Shida, T. K. F., de Oliveira, C. E. N., da Silva Frago de Campos, D., Los Angeles, E., Bernardo, C., dos Santos de Oliveira, L., et al. (2023b). Effect of freezing of gait and dopaminergic medication in the biomechanics of lower limbs in the gait of patients with Parkinson's disease compared to neurologically healthy. *Neuroscience Letters* 806, 137250. doi:10.1016/j.neulet.2023.137250
- Simoneau, G. G. (2010). Kinesiology of Waking. In *Kinesiology of the Musculoskeletal System: Foundations for Rehabilitation* (MOSBY Elsevier). 2nd edn., 627–681
- Stolze, H., Klebe, S., Petersen, G., Raethjen, J., Wenzelburger, R., Witt, K., et al. (2002). Typical features of cerebellar ataxic gait. *Journal of Neurology Neurosurgery and Psychiatry* 73, 310–312. doi:10.1136/jnnp.73.3.310
- Sun, H., Ye, Q., and Xia, Y. (2024). Predicting freezing of gait in patients with Parkinson's disease by combination of Manually-Selected and deep learning features. *Biomedical Signal Processing and Control* 88, 105639. doi:10.1016/j.bspc.2023.105639
- Takens, F. (1981). Detecting strange attractors in turbulence. In *Lecture Notes in Mathematics* (Berlin, Heidelberg: Springer Berlin Heidelberg), vol. 898. 366–381. doi:10.1007/BFb0091924
- The MathWorks Inc. (2019). Short-time Fourier transform. <https://www.mathworks.com/help/signal/ref/stft.html>. [Accessed 29 January 2024]
- Toro, B., Nester, C., and Farren, P. (2003). A review of observational gait assessment in clinical practice. *Physiotherapy Theory and Practice* 19, 137–149. doi:10.1080/09593980307964
- Torrence, C. and Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society* 79, 61–78. doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

- Van Crielinge, T., Saeys, W., Truijen, S., Vereeck, L., Sloot, L. H., and Hallemans, A. (2023). A full-body motion capture gait dataset of 138 able-bodied adults across the life span and 50 stroke survivors. *Scientific Data* 10, 852. doi:10.1038/s41597-023-02767-y
- Verlekar, T. T., Soares, L. D., and Correia, P. L. (2018). Automatic classification of gait impairments using a markerless 2D video-based system. *Sensors (Switzerland)* 18, 1–16. doi:10.3390/s18092743
- Vlachos, I. and Kugiumtzis, D. (2008). State Space Reconstruction for Multivariate Time Series Prediction. *Nonlinear Phenomena in Complex Systems* 11
- Wallot, S. and Mønster, D. (2018). Calculation of Average Mutual Information (AMI) and False-Nearest Neighbors (FNN) for the Estimation of Embedding Parameters of Multidimensional Time Series in Matlab. *Frontiers in Psychology* 9, 1679. doi:10.3389/fpsyg.2018.01679
- Wang, L., Sun, Y., Li, Q., Liu, T., and Yi, J. (2020). Two Shank-Mounted IMUs-Based Gait Analysis and Classification for Neurological Disease Patients. *IEEE Robotics and Automation Letters* 5, 1970–1976. doi:10.1109/LRA.2020.2970656
- Wang, Y., Zou, Q., Tang, Y., Wang, Q., Ding, J., Wang, X., et al. (2022). SAIL: A Deep-Learning-Based System for Automatic Gait Assessment from TUG Videos. *IEEE Transactions on Human-Machine Systems* 52, 110–122. doi:10.1109/THMS.2021.3123232
- Webber, C. L. and Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology* 76, 965–973. doi:10.1152/jappl.1994.76.2.965
- Whittle, M. W. (2007). *Gait analysis: An introduction* (Butterworth-Heinemann Elsevier), 4th edn.
- Yang, H. and Liu, G. (2013). Self-organized topology of recurrence-based complex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 23. doi:10.1063/1.4829877
- Yusuf, S. I., Adeshina, S., and Boukar, M. M. (2019). Parameters for human gait analysis: A review. In *Proceedings of the 15th International Conference on Electronics, Computer and Computation (ICECCO)*. 31–34. doi:10.1109/ICECCO48375.2019.9043216
- Zago, M., Luzzago, M., Marangoni, T., De Cecco, M., Tarabini, M., and Galli, M. (2020). 3D Tracking of Human Motion Using Visual Skeletonization and Stereoscopic Vision. *Frontiers in Bioengineering and Biotechnology* 8. doi:10.3389/fbioe.2020.00181
- Zeng, W., Liu, F., Wang, Q., Wang, Y., Ma, L., and Zhang, Y. (2016). Parkinson's disease classification using gait analysis via deterministic learning. *Neuroscience Letters* 633, 268–278. doi:10.1016/j.neulet.2016.09.043
- Zgolli, F., Henni, K., Haddad, R., Mitiche, A., Ouakrim, Y., Hagemester, N., et al. (2018). Kinematic Data Clustering for Healthy Knee Gait Characterization. In *Proceedings of the 2018 IEEE Life Sciences Conference (LSC)*. 239–242. doi:10.1109/LSC.2018.8572119

- Zhang, Y., Yan, W., Yao, Y., Ahmed, J. B., Tan, Y., and Gu, D. (2020). Prediction of Freezing of Gait in Patients With Parkinson's Disease by Identifying Impaired Gait Patterns. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28, 591–600. doi:10.1109/TNSRE.2020.2969649
- Zhao, A., Qi, L., Li, J., Dong, J., and Yu, H. (2018). A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data. *Neurocomputing* 315, 1–8. doi:10.1016/j.neucom.2018.03.032
- Řezanková, H. (2018). Different approaches to the silhouette coefficient calculation in cluster evaluation. In *Proceedings of the 21st International Scientific Conference AMSE*. 1–10

Appendix A

Experimental Protocol

Experimental setup:

- Treadmill gait trials:
 - Tripod height: 80 cm
 - Tripod distance to the treadmill: 60 cm
 - Camera frame rate: 60 fpm
 - Camera resolution: 2.7K
 - Len setup: Linear
 - Speed of the treadmill: 4.5 km/h
 - Duration of data acquisition: 90 s.
- Overground gait trials:
 - Length of the walking path: 10 m

Collected data:

- Subject age
- Subject weight
- Subject height
- Subject foot length
- Length from hip to knee marker
- Position in time of the hip marker
- Position in time of the knee marker
- Position in time of the ankle marker

Conducted gait trials:

- Treadmill gait trial without added heel
- Overground gait trial without added heel
- Treadmill gait trial with added heel (height of 1.5 cm) under the sole of the left ankle.
- Overground gait trial with added heel (height of 1.5 cm) under the sole of the left ankle.
- Treadmill gait trial with added heel (height of 2.5 cm) under the sole of the left ankle.
- Overground gait trial with added heel (height of 2.5 cm) under the sole of the left ankle.
- Treadmill gait trial with added heel (height of 3.5 cm) under the sole of the left ankle.
- Overground gait trial with added heel (height of 3.5 cm) under the sole of the left ankle.

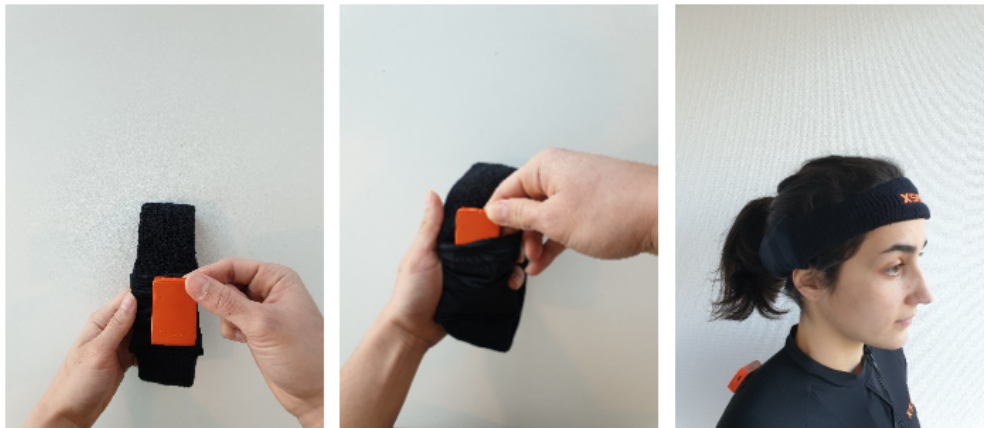
A.1 Sensor placement in MVN Awinda System

The MVN Awinda system utilizes wireless trackers attached to the body using a t-shirt and a set of straps. The following instructions outline the placement for one side of the human body. On the other side of the body, it is important to position the sensors symmetrically.

Please ensure that when placing the sensors on the other side, you mirror the positioning relative to the body's midline to maintain symmetry and accuracy in tracking.

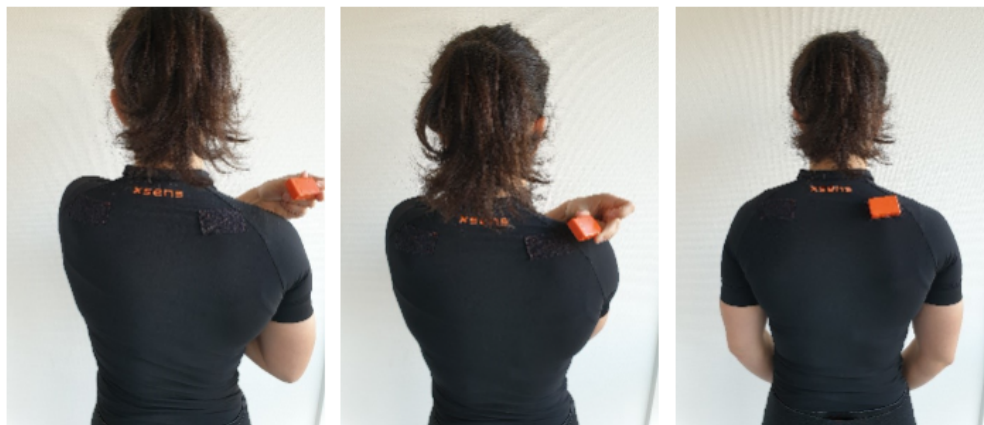
A.1.1 Head sensor

Place the sensor inside the headband and position it around your head. Ensure that the sensor's longer edge aligns with the band to maintain stability. The orientation of the headband on the head is not important, as long as the sensor is stable.



A.1.2 Shoulder sensors

Place the sensors in the Velcro patches located at the back of the t-shirt. To ensure optimal attachment of the sensor to the shirt, place it horizontally, allowing the entire sensor surface to be in contact with the Velcro patch.



A.1.3 Upper arm sensors

Place the sensor on the lateral side of the upper arm, between the muscle groups of the biceps and triceps.



A.1.4 Forearm sensors:

Place the sensor in close proximity to the wrist, on the frontal side.



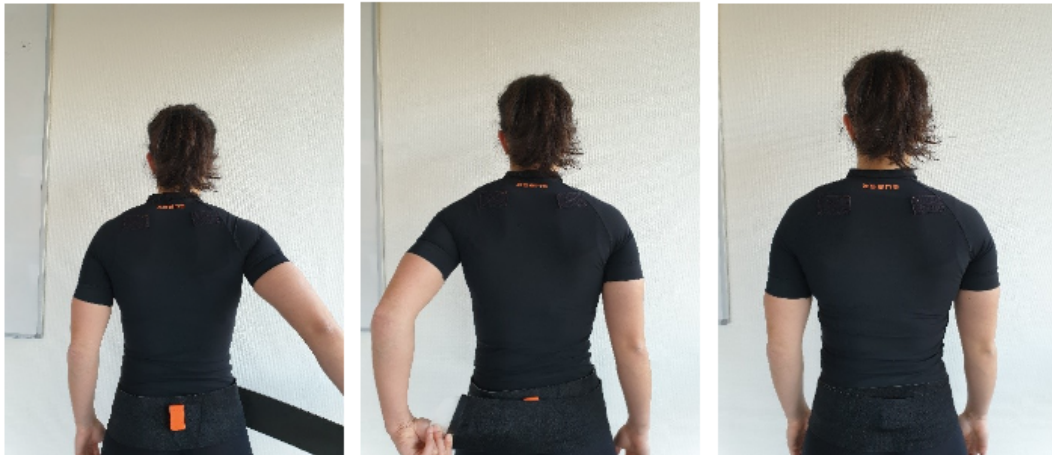
A.1.5 Hand sensors:

Place the sensor vertically in the pocket of the glove to ensure a more secure and comfortable fit. Each glove has two letter markings. The lateral mark (orange circle) is the size of the glove, and the upper marking (blue circle) is the hand (left or right) of the glove.



A.1.6 Pelvis sensor:

Use the largest strap provided and tighten it around the pelvis bone at the height of the anterior superior iliac spine. This ensures a secure fit. Place the sensor on the first layer of the strap. If possible, position it underneath the second layer of the strap for added stability. Place the sensor vertically so that the "XSSENS" label can be read right side up. Another reference point is to have the micro-USB port and power button of the sensor facing directly downward.



Please note that before calibration, there generally is not a specific orientation requirement for the sensors in the Awinda system, except for the pelvis sensor.

A.1.7 Sternum sensor:

Insert the sensor into the pocket located on the front of the t-shirt, ensuring that it is vertically placed.



A.1.8 Upper leg sensors:

Place the sensor on the middle of the lateral thigh, as this area is less likely to have fatty tissue and helps prevent the strap from sliding down. Take care not to leave the Velcro tip on the inner part of the leg, as it can potentially scratch the other leg during movement, leading to loosening of the strap.



A.1.9 Lower leg sensors:

Place the sensor on the front side of the tibia, near the knee.



A.1.10 Foot sensors:

Place the sensor on a strap and wrap the strap around the foot. Place the sensor on the upper front part of the foot, between the ankle joint and the toe joint.



A.2 Marker placement for vision-based method

The following instructions outline the placement for one side of the human body. On the other side of the body, it is important to position the sensors symmetrically. Please ensure that when placing the sensors on the other side, you mirror the positioning relative to the body's midline to maintain symmetry and accuracy in tracking.

A.2.1 Hip joint marker

With the subject's leg in an extended position, palpate to identify the femoral head. Place the marker directly at this location.

A.2.2 Knee joint marker

Ask the subject under study to flex his knee. Locate the gap between the femur and the tibia by palpating just below and behind the patella (kneecap). Then, palpate from the back of the knee to identify the back of the joint. Place the marker in the center of the bone between your finger and thumb to ensure accurate placement.

A.2.3 Ankle: Lateral Maleolus marker

Place your thumb and finger on either side of the ankle's bony protrusion. Locate the space between the toes and position the marker precisely in the center.

A.2.4 Foot metatarsal head marker

Palpate the outer side of the subject's foot to identify the most prominent bony protrusion in that region. Once located, place the marker directly on that point for accurate positioning.

Appendix B

Application of ICR-LLS methodology in the diagnosis of PD

B.1 Computed dissimilarity matrices

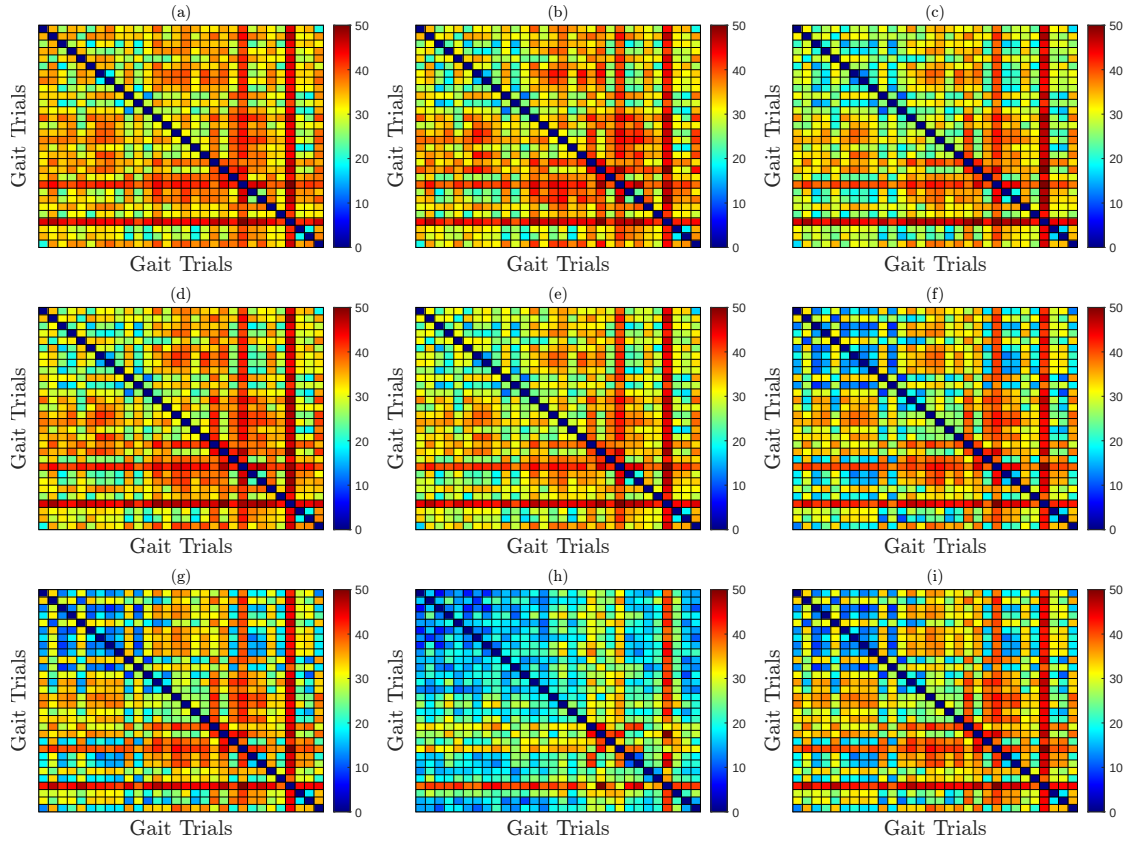


Fig. B.1: Dissimilarity matrices calculated from pairwise comparisons of the shank ICR trajectories of subjects with and without PD using (a) Euclidean distance, (b) mean dissimilarity, (c) RMS dissimilarity, (d) peak dissimilarity, (e) Manhattan distance, (f) DTW, (g) inverse cross-correlation, (h) cross-correlation + DTW, and (i) Förstner distance. In each matrix, ordered from top to bottom and from left to right, there are gait trials corresponding to 15 healthy subjects and 15 patients with PD, respectively.

B.2 Cluster analysis

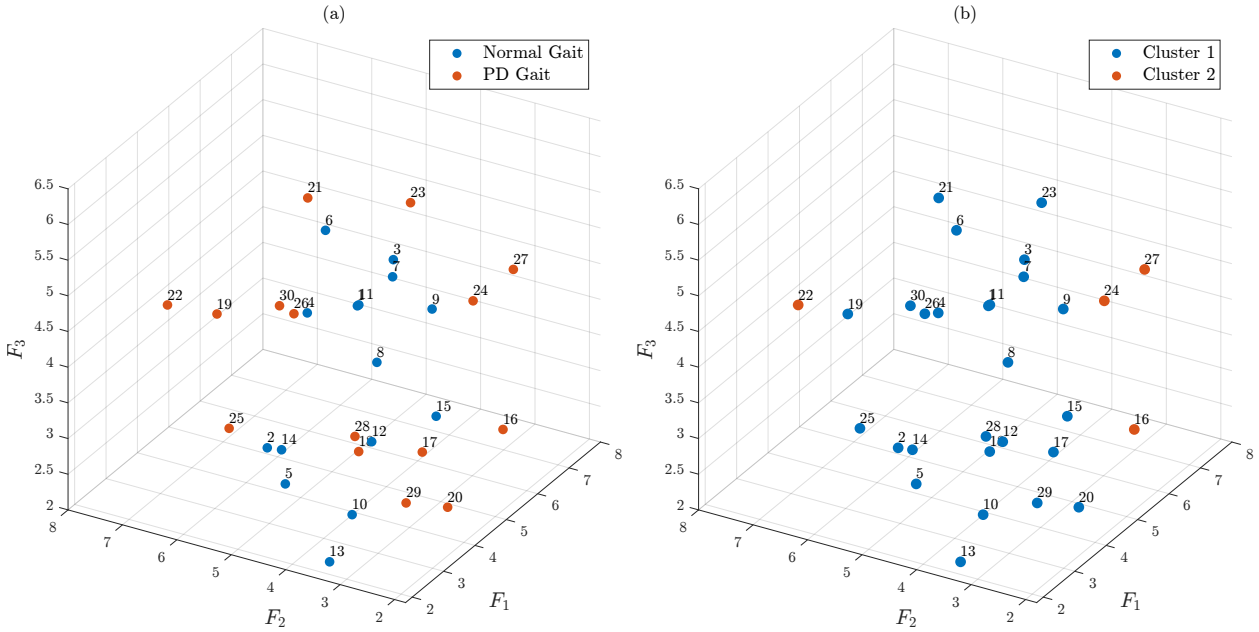


Fig. B.2: Clusters identified from the Euclidean distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.79$ and $MinPts = 7$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

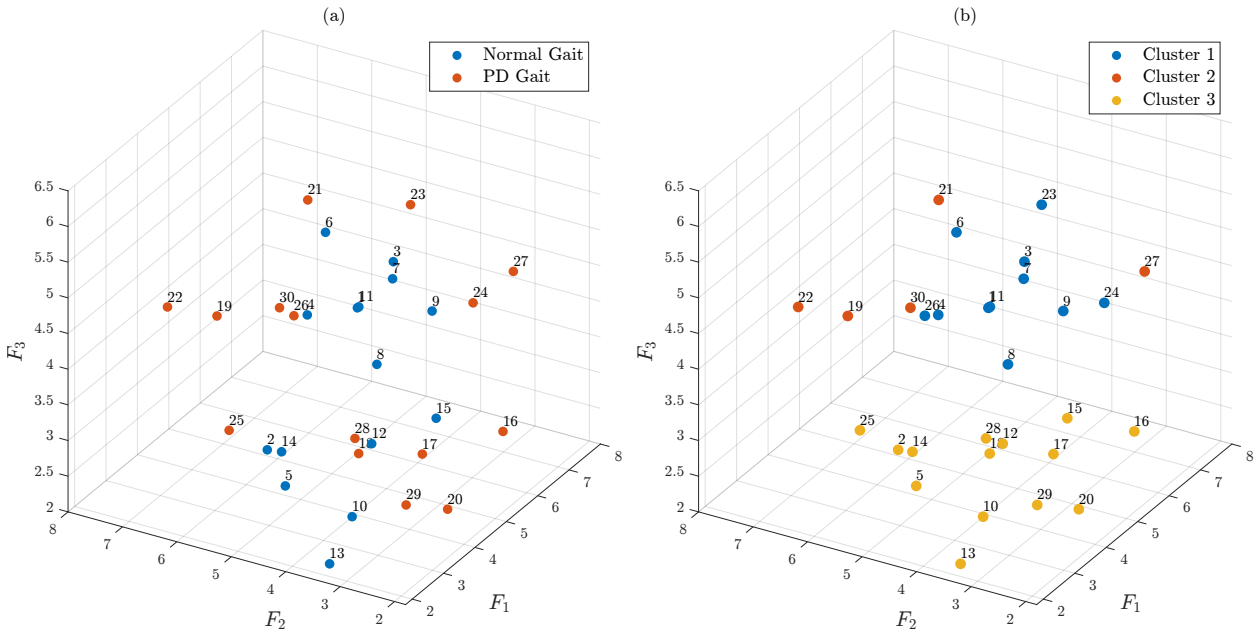


Fig. B.3: Clusters identified from the Euclidean distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.61$ and $MinPts = 5$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

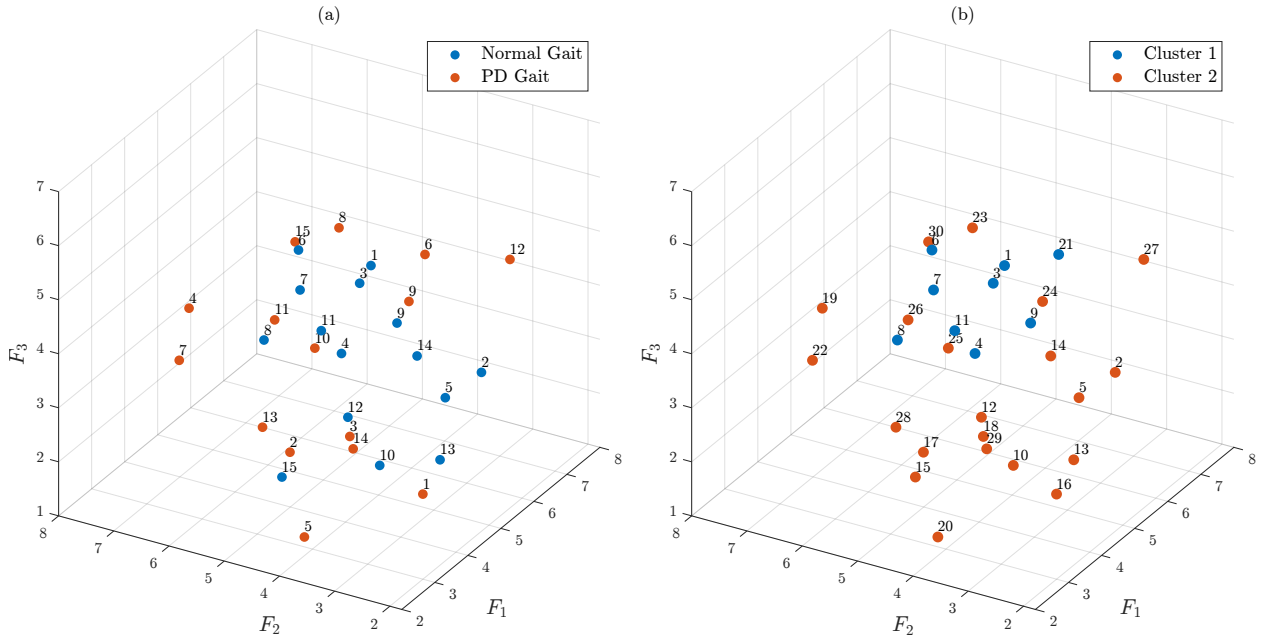


Fig. B.4: Clusters identified from the Mean dissimilarity-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.35$ and $MinPts = 5$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

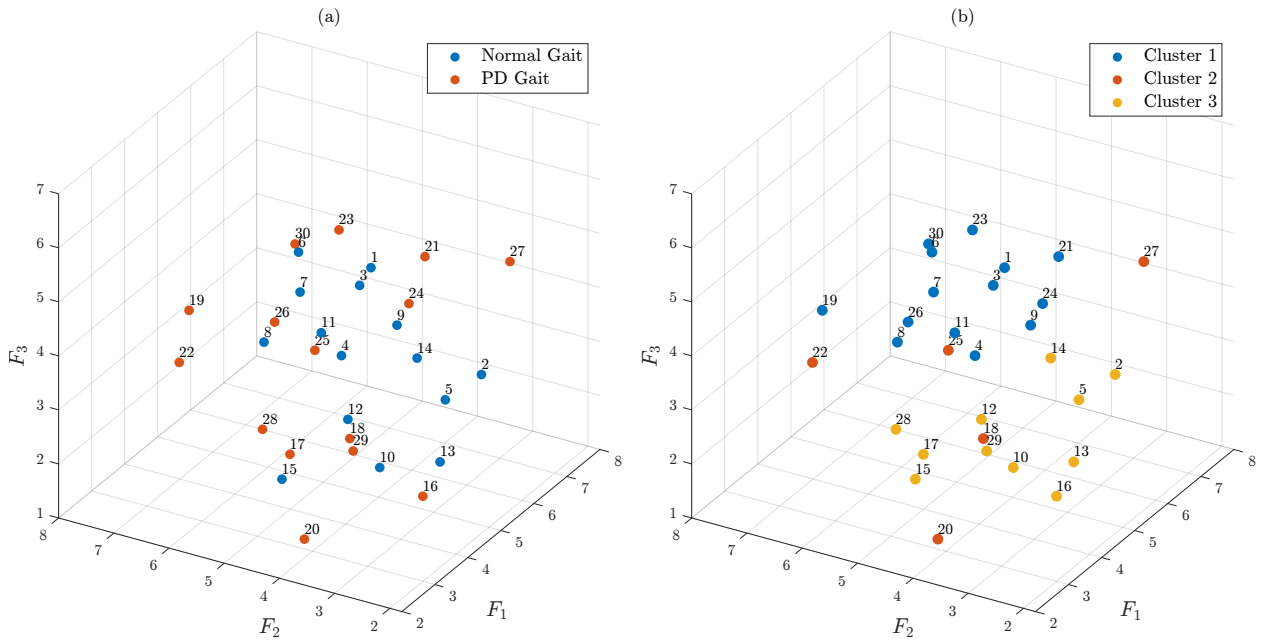


Fig. B.5: Clusters identified from the Mean dissimilarity-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.52$ and $MinPts = 3$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

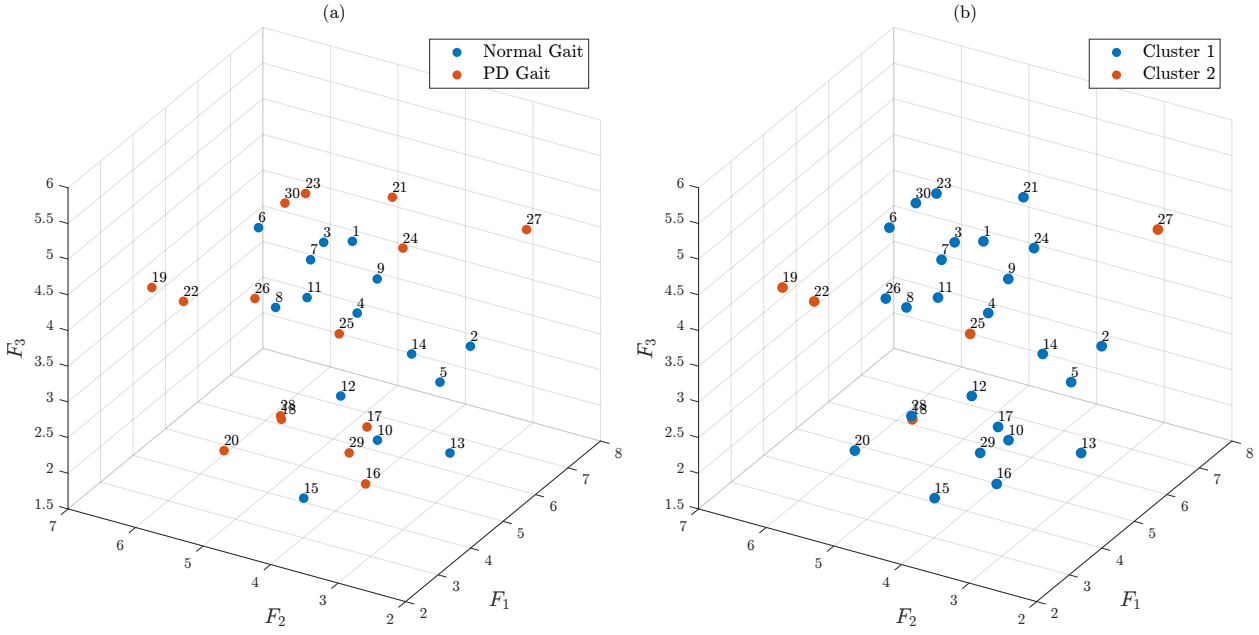


Fig. B.6: Clusters identified from the RMS dissimilarity-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.47$ and $MinPts = 6$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

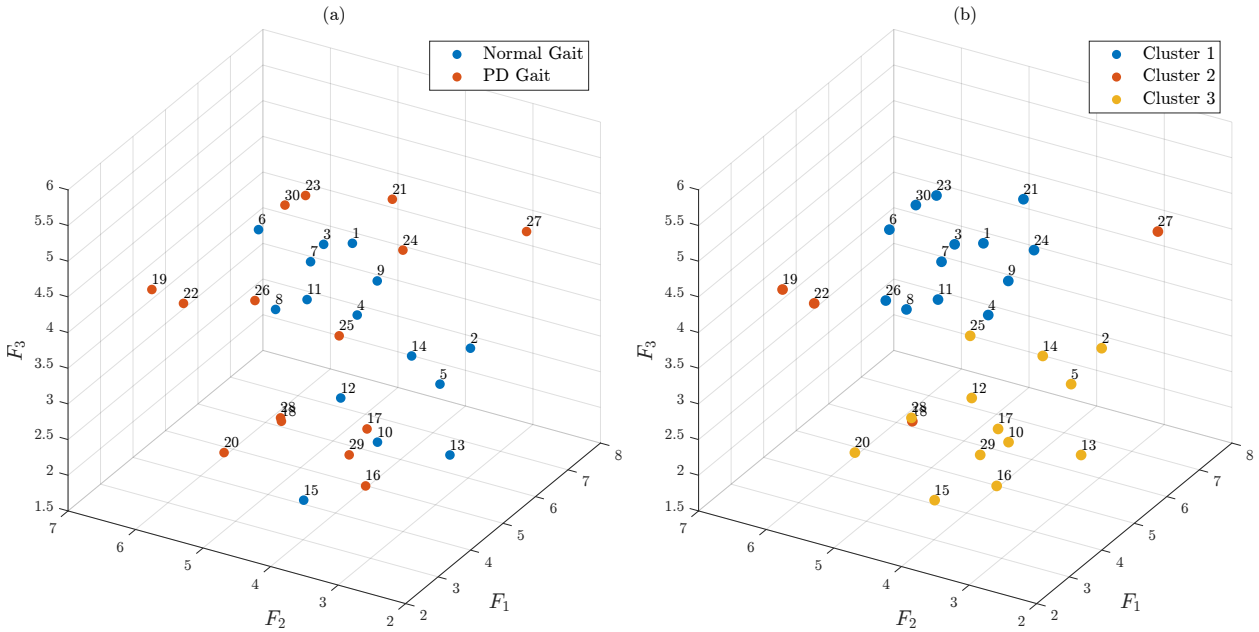


Fig. B.7: Clusters identified from the RMS dissimilarity-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.29$ and $MinPts = 3$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

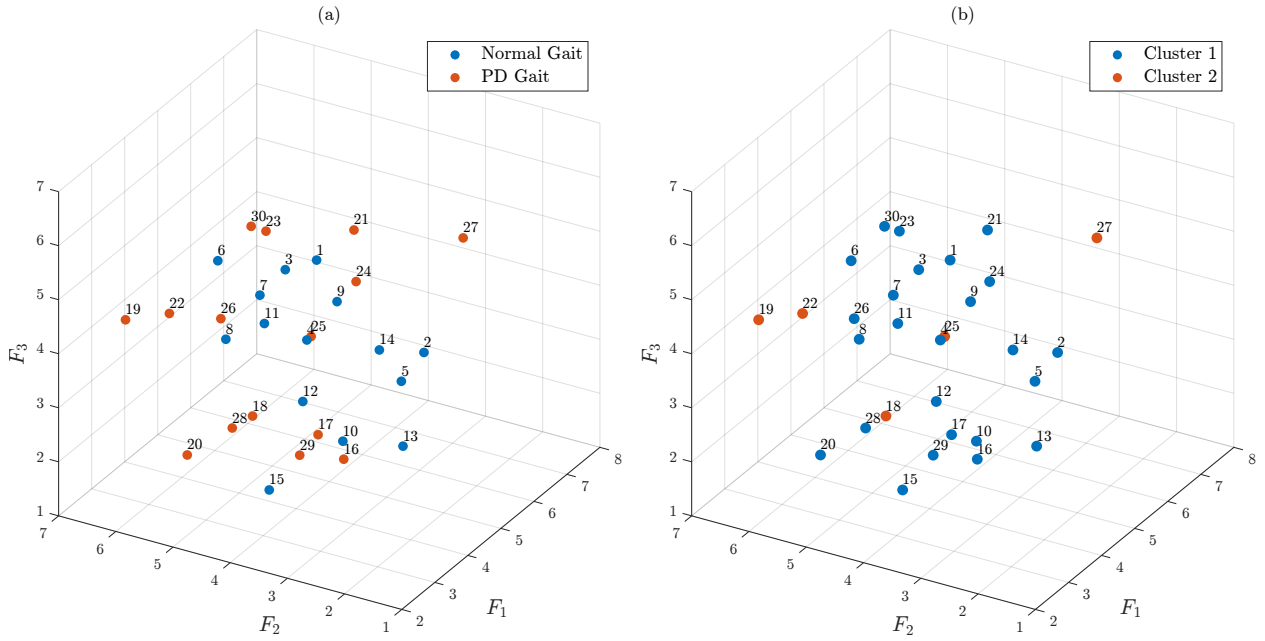


Fig. B.8: Clusters identified from the Peak dissimilarity-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.61$ and $MinPts = 6$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

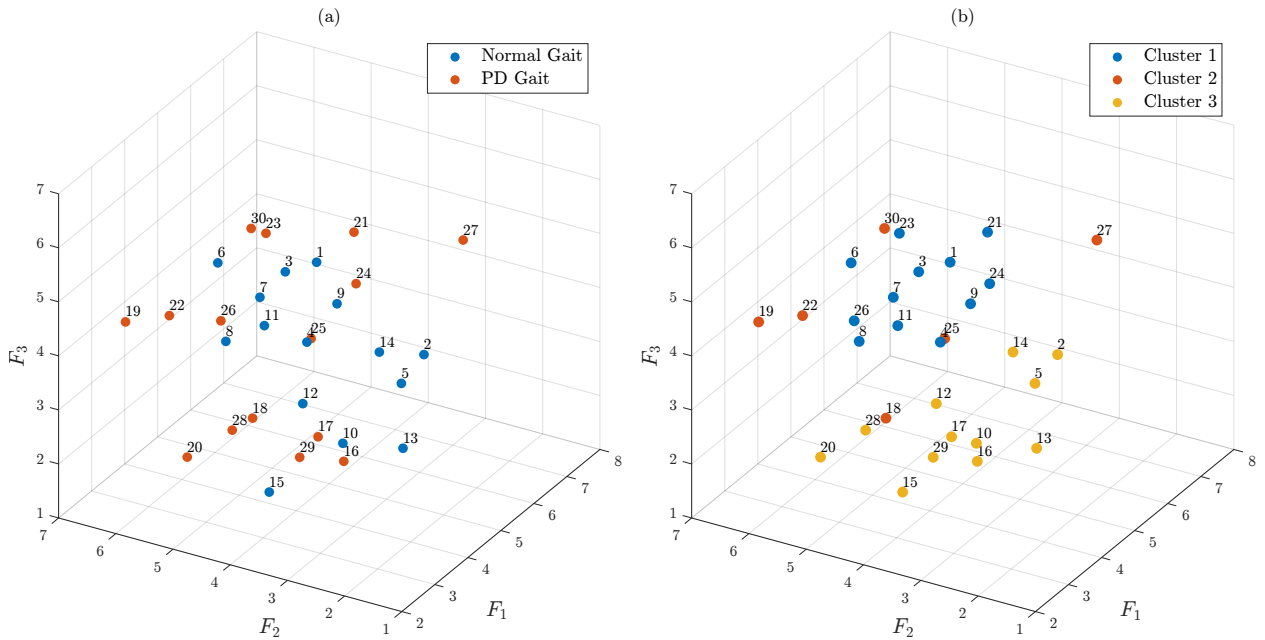


Fig. B.9: Clusters identified from the Peak dissimilarity-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.42$ and $MinPts = 4$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

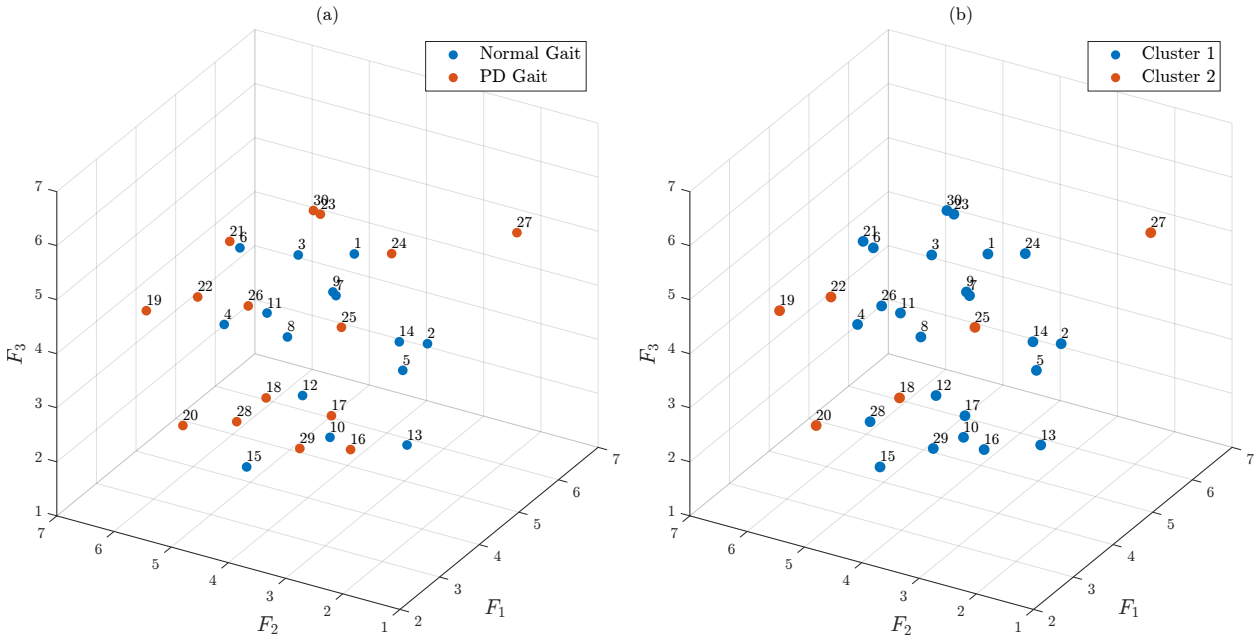


Fig. B.10: Clusters identified from the Manhattan distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.59$ and $MinPts = 6$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

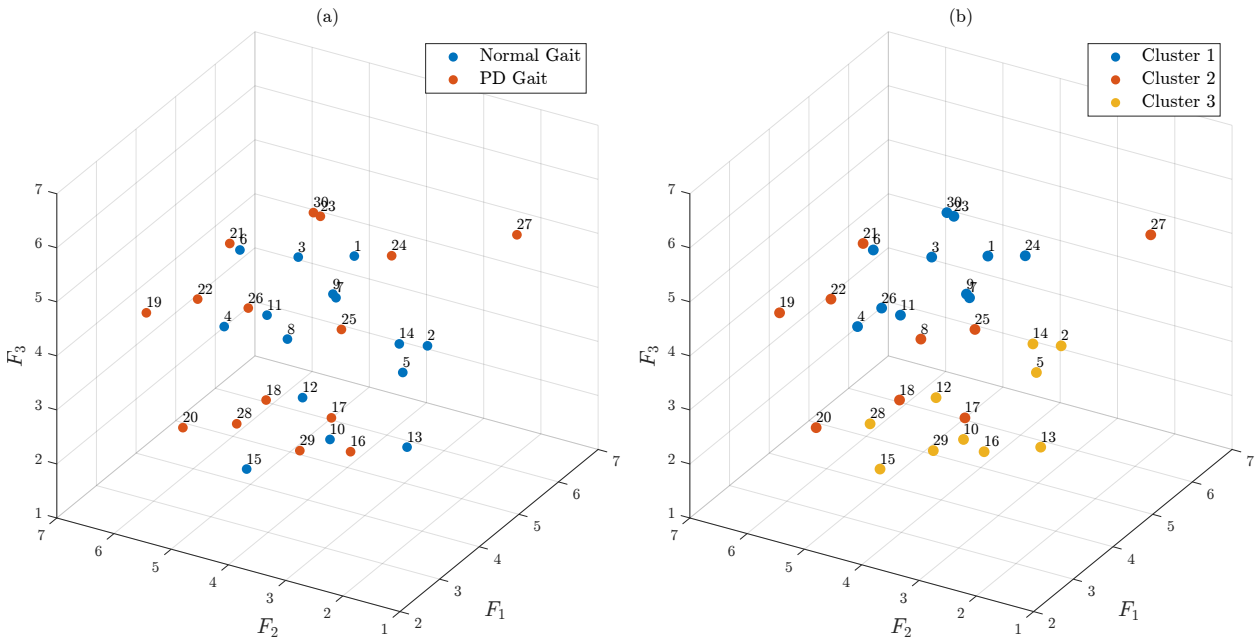


Fig. B.11: Clusters identified from the Manhattan distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.47$ and $MinPts = 6$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

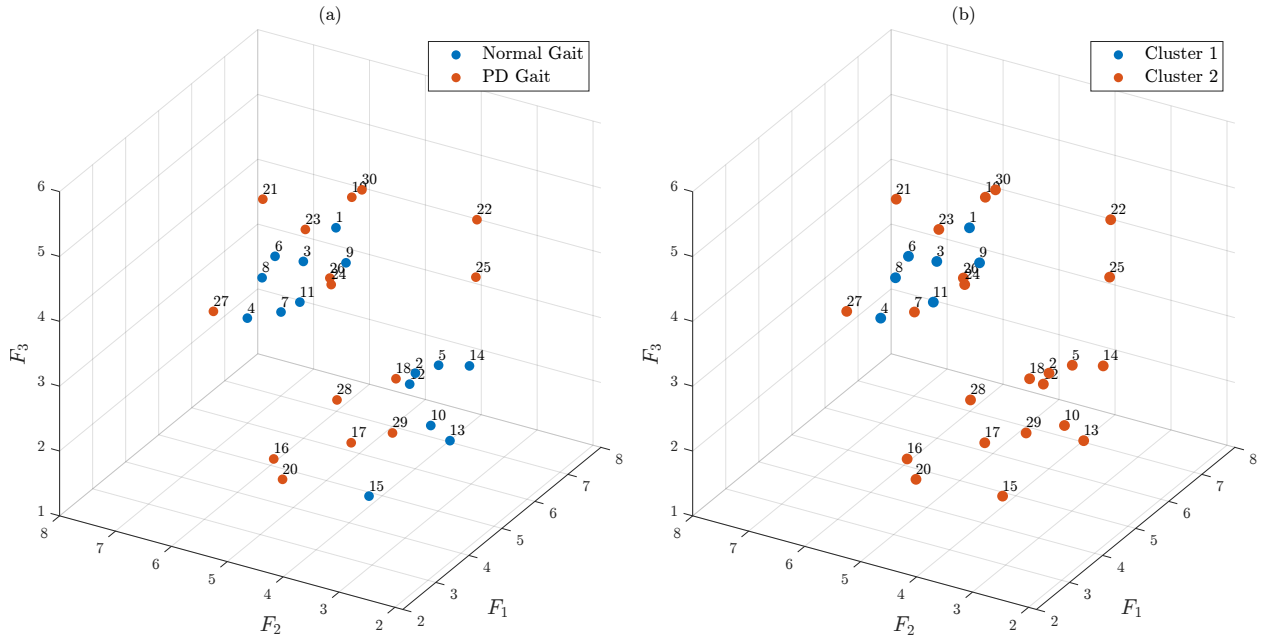


Fig. B.12: Clusters identified from the DTW-based network layout using the DBSCAN algorithm with parameters $\epsilon = 0.92$ and $MinPts = 5$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

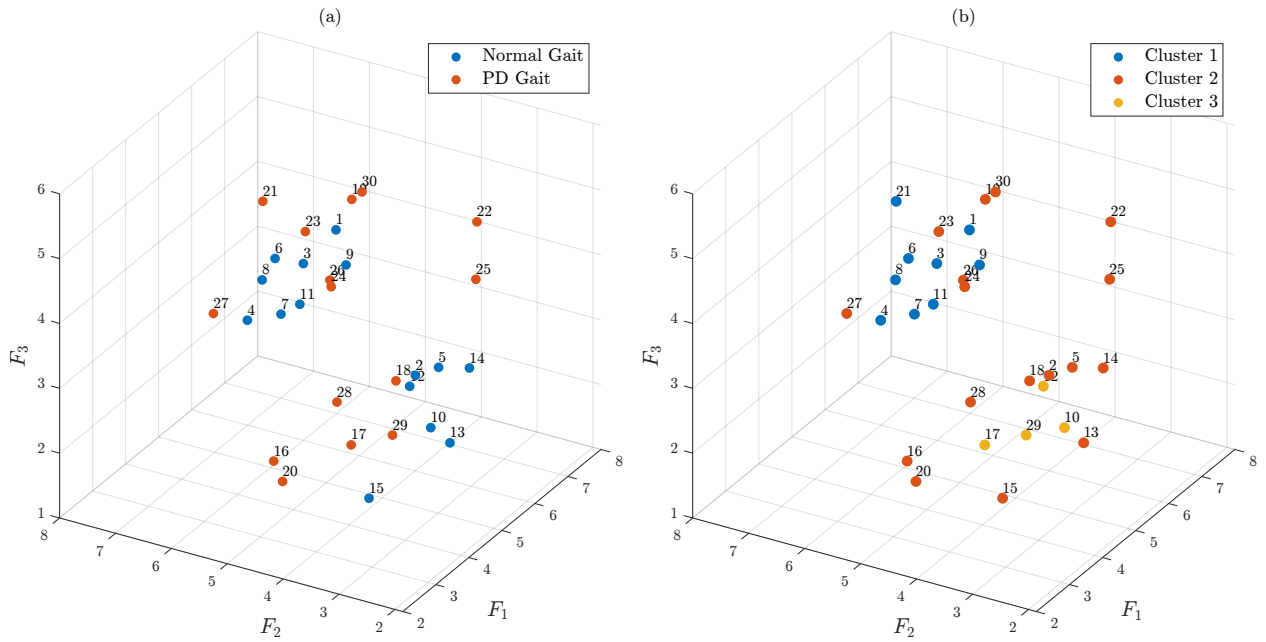


Fig. B.13: Clusters identified from the DTW-based network layout using the DBSCAN algorithm with parameters $\epsilon = 0.91$ and $MinPts = 4$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

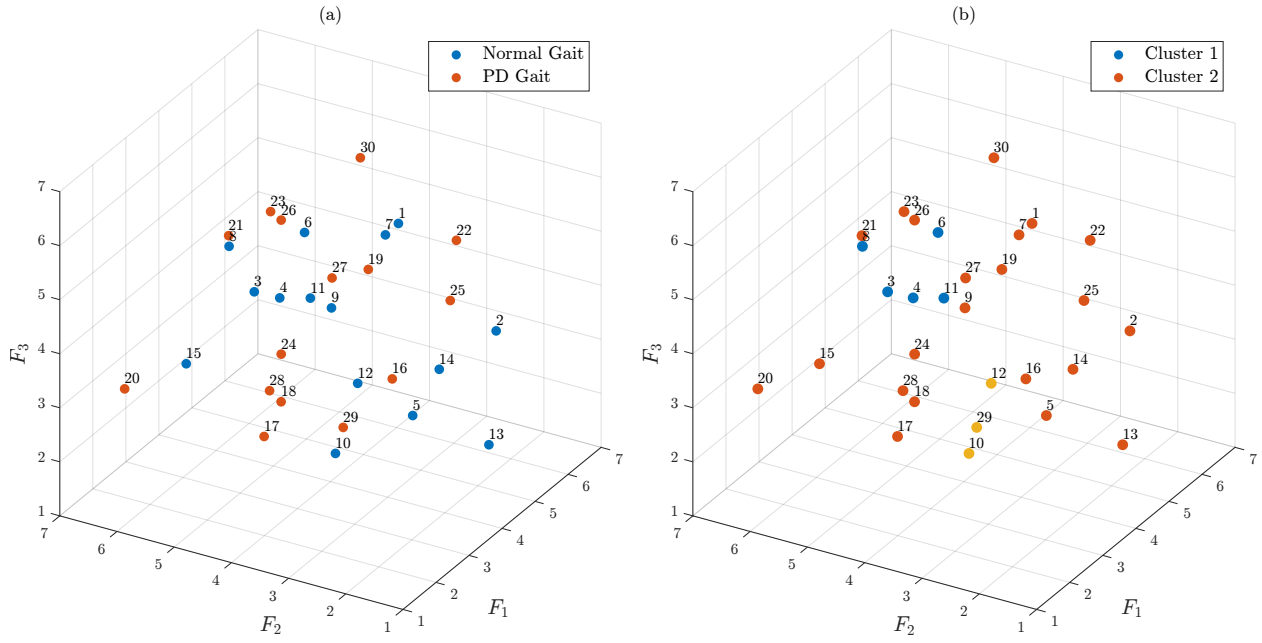


Fig. B.14: Clusters identified from the Cross-Correlation-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.22$ and $MinPts = 3$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

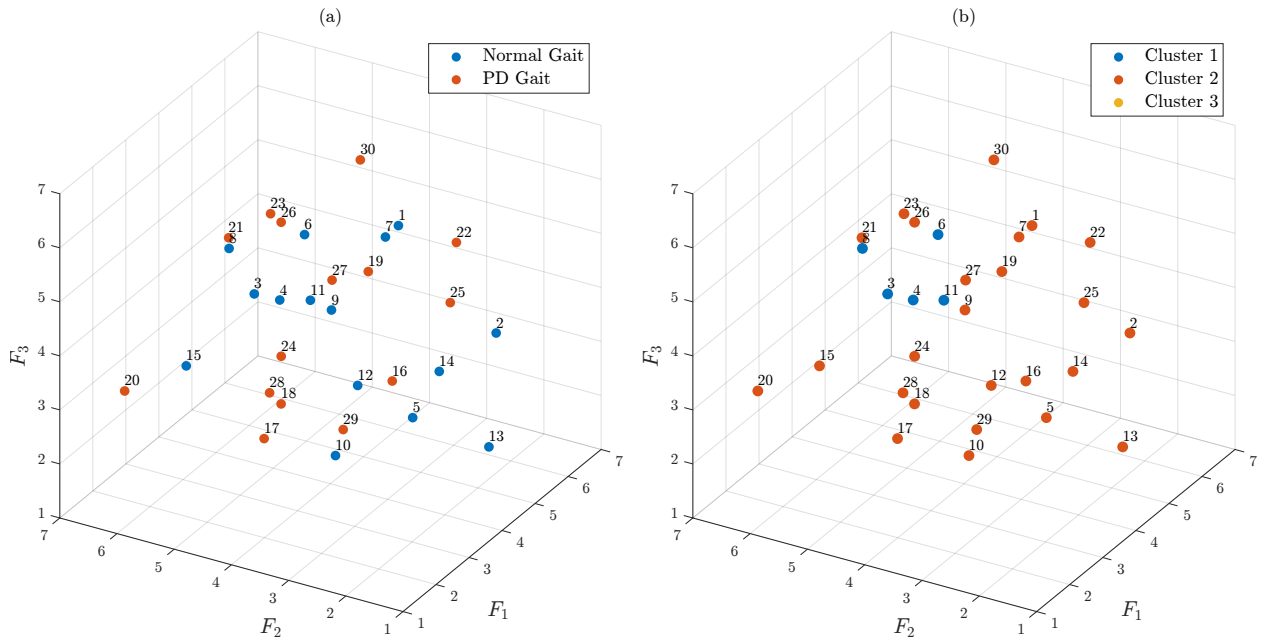


Fig. B.15: Clusters identified from the Cross-Correlation-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.07$ and $MinPts = 3$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

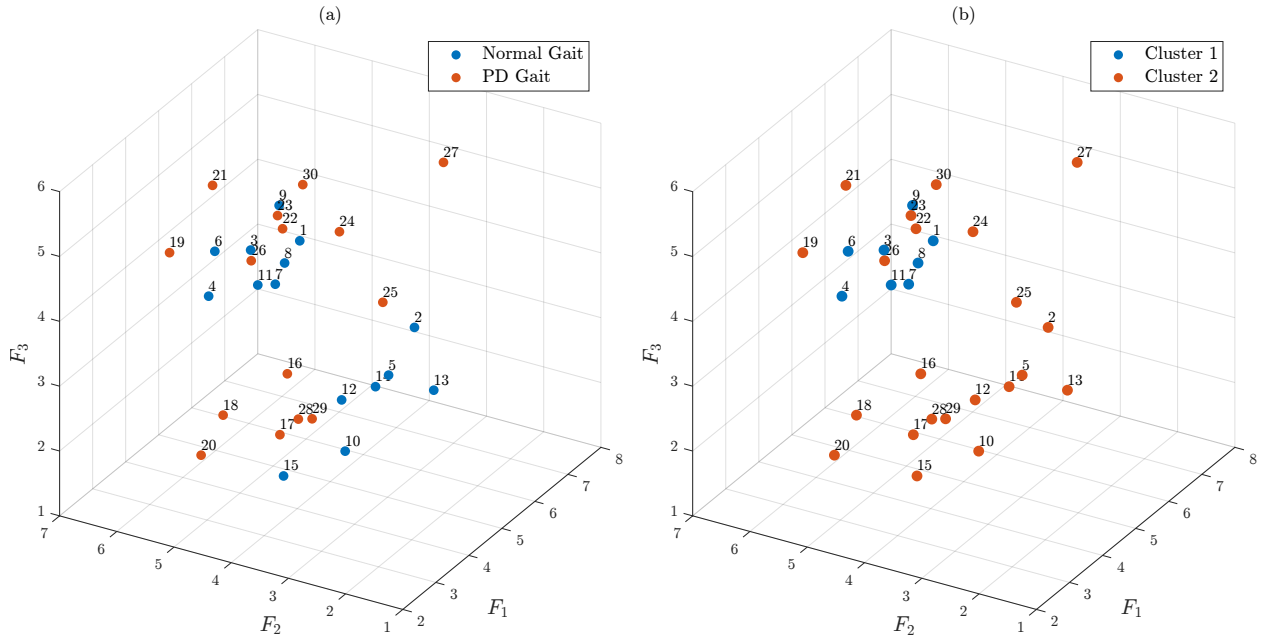


Fig. B.16: Clusters identified from the Cross-Correlation + DTW-based network layout using the DBSCAN algorithm with parameters $\epsilon = 0.89$ and $MinPts = 4$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

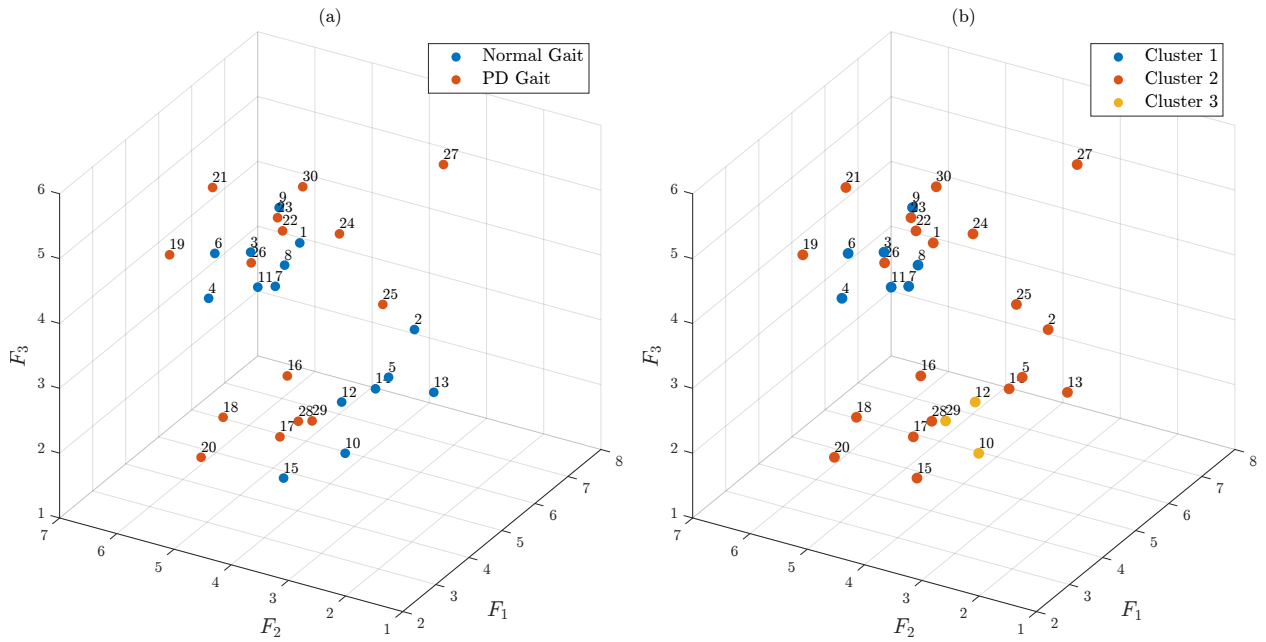


Fig. B.17: Clusters identified from the Cross-Correlation + DTW-based network layout using the DBSCAN algorithm with parameters $\epsilon = 0.85$ and $MinPts = 3$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

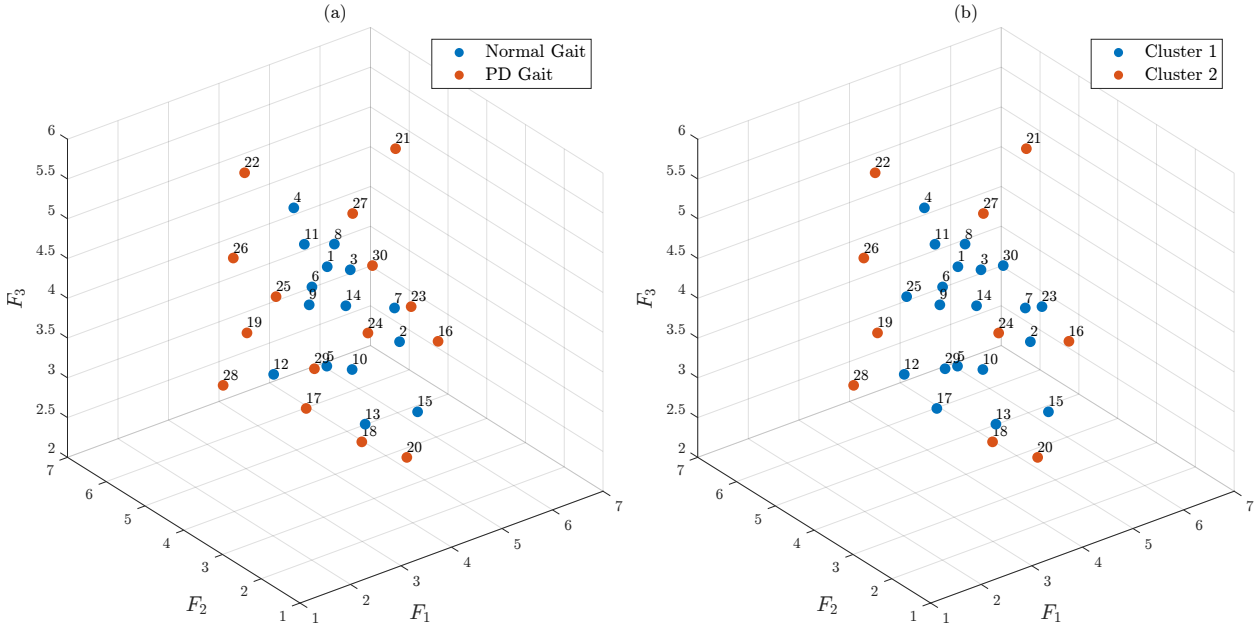


Fig. B.18: Clusters identified from the Förstner distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 1.04$ and $MinPts = 5$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

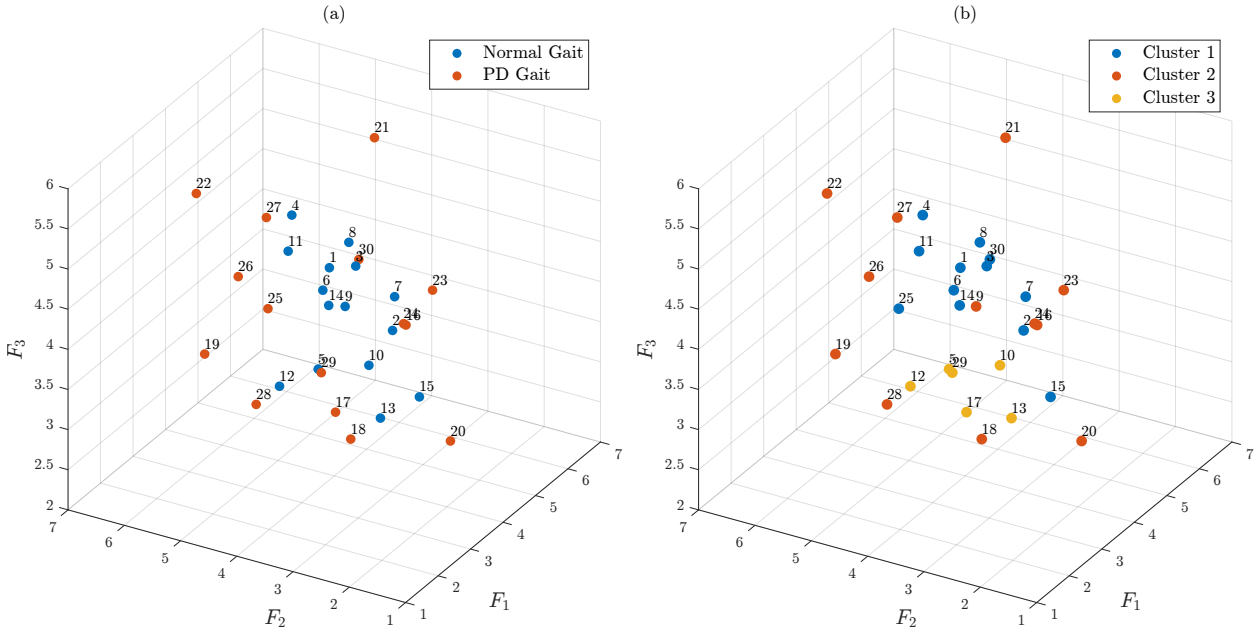


Fig. B.19: Clusters identified from the Förstner distance-based network layout using the DBSCAN algorithm with parameters $\epsilon = 0.99$ and $MinPts = 4$. (a) Results of the self-organization stage, where the blue dots are subjects with normal gait and the red dots denote patients with PD. (b) Identified clusters.

Kraków, dnia

.....
Imię i nazwisko autora rozprawy

.....
Imię i nazwisko promotora rozprawy

.....
Nazwa Wydziału

.....
Nazwa Katedry/Instytutu

OŚWIADCZENIE

Oświadczam, świadomy odpowiedzialności karnej za poświadczenie nieprawdy, że niniejszą rozprawę doktorską wykonałem osobiście i samodzielnie i że nie korzystałem ze źródeł innych niż wymienione w pracy.

Podpis

.....