



ScuDo
Scuola di Dottorato - Doctoral School
WHAT YOU ARE, TAKES YOU FAR



**UNIVERSITÀ
DEGLI STUDI
DI TORINO**

Doctoral Dissertation
Doctoral Program in Bioengineering and Medical-Surgical Sciences (32th Cycle)

Machine Learning approaches applied to the analysis of muscle activation patterns

Cristina Castagneri

* * * * *

Supervisors

Prof. Marco Knaflitz, Supervisor
Prof. Gabriella Balestra, Co-Supervisor

Politecnico di Torino
April, 2020

This thesis is licensed under a Creative Commons License, Attribution - Non-commercial - NoDerivative Works 4.0 International: see www.creativecommons.org. The text may be reproduced for non-commercial purposes, provided that credit is given to the original author.

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Cristina Castagneri
Turin, 05 April 2020

Summary

Gait analysis is the study of human locomotion and is widely used to assess normal and pathological functions of human walking. Particularly, the instrumented gait analysis allows for collecting signals from several sensors (i.e. basographic and electromyographic sensors) placed on the subject body, and to obtain a fully and completely operator-independent statistical analysis of the gait. However, the aggregation and the correct interpretation of data collected is a crucial point to provide an objective and correct assessment of the characteristics of human walking, usable for clinical practice.

In a previous work, a method based on a fuzzy logic classifier, named GAITSCORE, was developed to aggregate data collected from basographic sensors: the result of the system is a gait impairment score, that provide an objective and quantitative measurement of gait impairment. A further improvement of this system would be the integration of the information related to the timing of muscular activation during gait. However, human locomotion is characterized by a large intra-subject variability in muscle activation patterns: muscle activation onsets and offsets markedly vary from stride-to-stride also in individuals without neurological or orthopedic disorders. Therefore, before being able to correctly use the information extracted from the EMG signals, it is necessary to develop strategies to identify the variability in the muscle activation patterns and to provide tools for a correct interpretation of the acquired data. Particularly, for a correct interpretation of EMG data, a promising strategy consists in grouping strides characterized by similar EMG activation patterns.

The CIMAP (Clustering for Identification of Muscle Activation Patterns) is the method, based on machine learning techniques, that has been developed for this purpose: it allows to group strides in clusters sharing similar EMG activation patterns. Once the clusters are obtained, an element that is characteristic of each cluster

is defined. This element is the prototype and it represents all the elements of a specific cluster.

The first part of the thesis presents the CIMAP method, its optimization and the final validation. The CIMAP is applied to EMG signals acquired from different muscles of lower limb of both healthy and pathological subjects. The results obtained have proved that this method definitively improves the correct interpretation of the acquired signals. At first the method provides an organized representation of the most common type of muscle activation patterns of the individual, that may be useful to clinicians to simply identify possible criticism in subject walking. Then, the characterization of the all the strides belonging to a specific cluster with a single element (the prototype) allows for introducing the concept of principal activation. From the biomechanical point of view, the principal activations represent those activations necessary for accomplishing a specific motor task and they describe the essential contributions of a specific muscle to the movement. In practice, the principal activations of each muscle are extracted as the intersection of the corresponding cluster prototypes, obtained using CIMAP. The introduction of the concept of principal activations allows characterizing a subject with a single muscle activation pattern, representative of the specific muscle. This characterization allows for performing complex analysis of entire group of subjects.

To illustrate the potentialities of the principal activation concept, a first study is performed on a population of 100 school-age children. The results show that, analysing gait data using principal activations only, it is possible to clearly understand the biomechanical contribution of the analyzed muscles to the movement and to describe muscle activity in an simple form that may be useful in clinical practice, for a correct interpretation of gait data. As an example, in this study, it is possible to identify and describe phenomena related to gait maturation.

It is followed by two studies aimed to compare muscle activation in healthy and pathological subjects, using principal activations. In the first study gait data acquired from a group of patients effected by normal pressure hidrocephalus are compared with gait data acquired from a group of age-matched healthy subjects (controls). In the second study, gait data acquired from 20 patients with Total Hip Arthroplasty at 3, 6 and 12 months after surgery and 20 age-matched controls are analysed and compared. The dataset comparison has always been a very challenging problem due to the EMG intra-subject variability. The results obtained in these studies highlight how the application of the CIMAP algorithm and the principal activation extraction may be powerful tools for the aggregation of EMG results of huge dataset, and they may allow a clear comparison between different datasets.

Then, two studies are presented in order to introduce tools, based on muscle principal activations, for the quantification and interpretation of specific aspects of gait. In the first study an EMG asymmetry index, for assessing muscle-activation asymmetry in cyclic movements, is introduced and validated over a population of 114 subjects consisting of healthy subjects and both neurological and orthopaedic patients. Different asymmetry levels are expected to be found on each group, considering the different disorders and treatments which patients underwent. The value obtained for the asymmetry index are consistent with the expected asymmetry level of each specific group and this suggests that the index can be successfully used in clinics for an objective assessment of the asymmetry of muscle activation patterns during locomotion.

In the second study, two indices are presented. The first one is a muscle-specific functionality index that quantifies the similarity of the activation pattern of a specific muscle of a subject with that of the corresponding muscle of a healthy population. The second one, considering a pool of muscles, is a global index to quantify the distance between the functionality of a specific subject and that of a reference population. The effectiveness of the muscle-specific and global indices is validated by applying these indices to a group of 25 healthy children and to a group of 25 hemiplegic children to measure the distance of specific subjects and of the entire populations from a reference population of 55 healthy children. The results obtained show that these indices may be useful in clinics: it may be used for providing an overall evaluation of muscle functionality during both the first instrumental examination of a subject and when a subject is evaluated successively, along a rehabilitation program.

A brief introduction to the application of principal activation extraction to muscle synergy analysis is then provided. The results suggest that this kind of processing provide a more interpretable assessment of the modular organization of the central nervous system during a walking task without any loss of information.

In conclusion, this thesis describes several tools that have been developed and validated to analyse the muscle activation characteristics during gait. They may be useful in clinical practice (to help clinicians in the interpretation of data of single individuals or to assess patient improvement during a rehabilitation protocol) and in clinical studies (to characterize gait anomalies correlated to a specific pathology). Moreover, they may be used in research studies as demonstrated by the muscle synergies application.

Finally, a future work will be the integration of the indices in GAITSCORE to construct an initial dashboard for analysing the gait of a subject.

Acknowledgment

First and foremost, I would like to thank my supervisors, prof. Gabriella Balestra and prof. Marco Knaflitz, for their guidance during these years.

I would also like to thank ing. Samanta Rosati and prof. Valentina Agostini for helping me during my PhD.

I would like to thank all my colleagues at Biolab; a special thank goes to Massimo and Marco, for their sincere friendship and for always encouraging me to achieve my personal and professional goals.

Last but not least, I would like to thank my friend Valerie, simply for always being there, and my beloved family, for their constant and loving support in every moment of my life.

*“All truly great
thoughts are conceived
while walking.”*

Friedrich Nietzsche

Contents

1. Introduction.....	1
1.1 Gait analysis.....	1
1.2 Instrumentation and signals in gait analysis	2
1.2.1 Basographic signal.....	2
1.2.2 Electromyographic signal	3
1.3 Statistical gait analysis.....	4
1.3.1 Segmentation and classification of gait cycle.....	6
1.3.2 Observation of muscle activations.....	7
1.4 STEP32: a movement analysis system	7
1.4.1 Sensors	8
1.4.2 Basographic and electromyographic analysis.....	9
2. CIMAP: Clustering for Identification of Muscle Activation Patterns.....	15
2.1 Introduction.....	15
2.2 Method	16
2.2.1 EMG signal pre-processing	16
2.2.2 Implementation of the CIMAP algorithm.....	17
Dataset preparation	17
Clustering of each dataset	18
2.3 Testing	19
2.4 Validation.....	22
3. Optimization of the CIMAP algorithm.....	25
3.1 Introduction.....	25
3.2 Tuning of the dendrogram parameters.....	25
3.2.1 Centroid definition.....	26
3.2.2 Linkage method	26
3.2.3 Distance metric	27

3.2.4	Selection of the cutoff point.....	27
3.3	Clustering evaluation	30
3.4	Testing	30
3.5	Validation.....	34
4.	Principal activation extraction and analysis.....	36
4.1	Introduction.....	36
4.2	Population characterization using PAs	38
4.3	Characterization of school-age children population	40
4.3.1	Population and gait data acquisition	40
4.3.2	Data analysis and statistical analysis	41
Behavior of the entire pediatric population	41	
Differences among age groups.....	42	
Ankle control	42	
4.3.3	Results.....	42
Behavior of the entire pediatric population	42	
Differences among age groups.....	43	
Ankle control	45	
4.3.4	Discussion.....	45
Behavior of the entire pediatric population	46	
Differences among age groups.....	47	
Ankle control	48	
4.3.5	Conclusions.....	49
4.4	Assessment of muscle activation differences between idiopathic normal pressure hydrocephalus patients and healthy controls.....	49
4.4.1	Population and gait data acquisition	50
4.4.2	Data analysis and statistical analysis	50
4.4.3	Results.....	51
4.4.4	Discussion.....	53
4.4.5	Conclusion	55
4.5	Longitudinal assessment of muscle function in patients with Total Hip Arthroplasty.....	55
4.5.1	Population and gait data acquisition	56
4.5.2	Data analysis	56
4.5.3	Results.....	57
4.5.4	Discussion.....	58

4.5.5	Conclusion	59
5.	EMG-based indices for gait data interpretation	60
5.1	Introduction.....	60
5.2	EMG Asymmetry Index.....	61
5.2.1	Population and gait data acquisition	62
5.2.2	Definition of the sEMG asymmetry index.....	63
5.2.3	Other EMG indices defined in literature.....	65
5.2.4	Statistical analysis.....	65
5.2.5	Results and discussion	66
Mega TKR patients and Healthy Adults.....		69
TKR patients and Healthy Elderly.....		69
THA patients and Healthy Elderly.....		69
Hemiplegic and Healthy Children		70
iNPH patients and Healthy Elderly.....		70
5.2.6	Comparison with indices found in literature.....	70
5.2.7	Conclusion	71
5.3	Muscle Functional Indices	71
5.3.1	Definition of the muscle functionality indices.....	72
Representation of the Muscle Function of the Reference Population.....		73
Calculation of the Muscle Functional Indices		76
5.3.2	Validation.....	78
Population and gait data acquisition		78
Signal pre-processing.....		80
Representation of the Muscle Function of the Reference Population.....		80
Calculation of the Muscle Functional Indices		80
Statistical Analysis.....		80
5.3.3	Results and discussion	80
5.3.4	Conclusion	85
6.	Principal and secondary activation extraction applied to muscle synergy analysis.....	86
6.1	Introduction.....	86
6.2	Gait data acquisition	87
6.3	Influence of PA extraction on muscle synergy analysis	88
6.3.1	Standard pre-processing.....	88
6.3.2	Pre-processing using principal activations	88

6.3.3	Muscle synergy extraction	88
6.3.4	Muscle synergy analysis	89
6.3.5	Results and discussion	90
6.4	Influence of SA extraction on muscle synergy analysis	92
6.4.1	Pre-processing using secondary activations	93
6.4.2	Results and discussion	93
6.4.3	Conclusion	94
7.	Conclusions.....	96
8.	References.....	100
9.	Appendix A: Dataset description.....	107
10.	Appendix B: Distance metric tests.....	111
11.	Appendix C: Cutoff point tests	122

List of Tables

Table 3.1. #OCC results (8 combination of parameters)	31
Table 3.2. #OCC results (4 combination of parameters)	31
Table 3.3. Comparison of Original and new CIMAP	32
Table 3.4. Comparison of Original and new CIMAP (Validation)	34
Table 4.1. Population details	40
Table 4.2. Comparison among age groups: results of the Hotelling's T-square test (p-values).....	45
Table 4.3. Population details	50
Table 4.4. Gait phase time duration.....	52
Table 4.5. Lilliefors test results	52
Table 4.6. Student t-test and Wilcoxon test results	52
Table 4.7. Population details	56
Table 5.1. Population details	64
Table 5.2. Comparison between pathological groups and age-matched controls	68
Table 5.3. Population details	79
Table 5.4. Lilliefors test results	81
Table 5.5. Student t-test results.....	81
Table A.1. Dataset details	110

List of Figures

Figure 1.1. Basographic levels.....	3
Figure 1.2. Example of EMG signal.....	4
Figure 1.3. Example of intra-subject variability in EMG signal	5
Figure 1.4. Example of basographic classification.....	6
Figure 1.5. Basographic sensors connected to preamplifier/decoder	8
Figure 1.6. EMG probes	9
Figure 1.7. Histograms of muscle activity modalities	10
Figure 1.8. Representation of EMG activation intervals (example 1).....	11
Figure 1.9. Representation of EMG activation intervals (example 2).....	13
Figure 1.10. Example of EMG pattern variability.....	14
Figure 2.1. CIMAP processing pipeline	17
Figure 2.2. Example of clustering result.....	21
Figure 2.3. Mean cluster SD	22
Figure 2.4. Mean cluster SD	23
Figure 2.5. Example of clustering result.....	24
Figure 3.1. Linkage Methods.....	26
Figure 3.2. Cutoff points.....	28
Figure 3.3. Comparison of three cut-off criteria.....	29
Figure 3.4. Comparison between CIMAP results (original and optimized version)	33
Figure 4.1. Example of principal activation extraction	37
Figure 4.2. Example of sensors positioning.....	39
Figure 4.3. Probes positioning.....	41
Figure 4.4. Activation representation of 100 school-aged children.....	43
Figure 4.5. Activation representation of three sub-population of the 100 school-aged children.....	44
Figure 4.6. Results found in [14]	46
Figure 4.7. Probes positioning	51
Figure 4.8. Activation representation of iNPH patients and controls.....	53
Figure 4.9. Probes positioning	57
Figure 4.10. Activation representation of THA patients at 3, 6 and 12 months after surgery, and controls	58
Figure 5.1. Overview of classification of symmetry assessment approaches [60]	61

Figure 5.2. Probes positioning	63
Figure 5.3. Mean values and standard errors of asymmetry indices	67
Figure 5.4. Pipeline for the definition of the muscle functional indices.....	72
Figure 5.5. Examples of penalty term application.....	74
Figure 5.6. Example of the outlier removal pipeline (Tibialis Anterior muscle, right side).....	75
Figure 5.7. Example of the construction and representation of the reference vector R (muscle: Tibialis Anterior).....	76
Figure 5.8. Example of probability density function estimated and J05 value computing	78
Figure 5.9. Probes positioning	79
Figure 5.10. Reference vector representation	82
Figure 5.11. Examples of Jm value representation.....	84
Figure 5.12. MFI values (mean value and standard deviation) over the populations of hemiplegic and healthy children.....	85
Figure 6.1. Probes positioning	87
Figure 6.2. Activation coefficients Ck and weight vectors Wk obtained with two different processing techniques	90
Figure 6.3. Example of secondary activation extraction	92
Figure 6.4. Example of completely auxiliary secondary activation extraction	93
Figure 6.5. Activation coefficients Ck and weight vectors Wk obtained with three different processing techniques	94
Figure 6.4. Example of completely auxiliary secondary activation extraction	93
Figure 6.5. Activation coefficients Ck and weight vectors Wk obtained with three different processing techniques	94
Figure B.1. Clustering results (Healthy male, 25 years old)	112
Figure B.2. Clustering results (Healthy male, 30 years old)	113
Figure B.3. Clustering results (Healthy female, 28 years old)	114
Figure B.4. Clustering results (Healthy male, 26 years old)	115
Figure B.5. Clustering results (THA patient, female, 70 years old).....	116
Figure B.6. Clustering results (THA patient, female, 61 years old).....	117
Figure B.7. Clustering results (Mega TKR patient, male, 20 years old)	118
Figure B.8. Clustering results (Mega TKR patient, male, 39 years old)	119
Figure B.9. Clustering results (TKR patient, female, 77 years old)	120

Figure B.10. Clustering results (TKR patient, female, 59 years old)	121
Figure C.1. Clustering results (iNPH patient, female, 85 years old)	123
Figure C.2. Clustering results (THA patient, female, 70 years old)	124
Figure C.3. Clustering results (THA patient, male, 65 years old)	125
Figure C.4. Clustering results (THA patient, female, 75 years old)	126
Figure C.5. Clustering results (Healthy male, 51 years old)	127
Figure C.6. Clustering results (Healthy male, 24 years old)	128
Figure C.7. Clustering results (Healthy female, 28 years old)	129
Figure C.8. Clustering results (Healthy male, 30 years old)	130
Figure C.9. Clustering results (Healthy female, 25 years old)	131
Figure C.10. Clustering results (iNPH patient female, 85 years old)	132

Chapter 1

Introduction

1.1 Gait analysis

Gait analysis is the study of human locomotion and is used to assess normal and pathological functions of human walking [1]. In clinical scenario, gait analysis is widely used to identify gait abnormalities [2], to assess residual walking ability in patient with neurological disorders [3], to define and assess rehabilitation protocol after orthopedic surgery [4] and to evaluate clinical outcomes [5].

Gait analysis can range from the simple observation of the subject during a walking session, to using fully computerized three-dimensional motion analysis systems [6], but, until recently, the ability of studying human gait was limited mainly to qualitative naked eye observation [7].

In 1970s, thanks to the availability of low-cost video cameras and computers, the stereophotogrammetric systems became a milestone in the story of human motion analysis, allowing a 3D reconstruction of human gait. At present, these systems are used in many motion analysis laboratories, but they require complex and expensive instrumentation, have high maintenance costs and are quite difficult to use in the clinical routine practice [8]. Moreover, stereophotogrammetric systems, even if based on 3D techniques, allow to obtain reliable measurements of joint angles only in the sagittal plane and in a limited volume of space.

Since 1980s, the instrumented gait analysis has emerged as a new powerful tool for human locomotion analysis, based on cheaper and simpler systems that do not required a settled laboratory environment to be operated. These tools are based on multichannel recording systems, connected through several sensors, on the subject

body. Signals coming from the sensors are acquired by means of a personal computer and processed using a dedicated software. These systems allow to collect signals from several sensors placed on the subject body, and to obtain a fully and completely operator-independent statistical analysis of gait. For these reasons, the instrumented gait analysis has revealed to be a powerful tool to provide quantitative, repeatable and reliable measure of gait parameters.

1.2 Instrumentation and signals in gait analysis

In gait analysis, measurement instruments commonly used include foot-switches and electromyographic sensors, for the acquisition of basographic and electromyographic signals, respectively.

1.2.1 Basographic signal

Foot-switches are sensors placed under the foot, in specific anatomical landmarks, aimed to reveal the foot-floor contact and to measure the corresponding temporal parameters. For example, using three sensors (one placed beneath the heel and the other two on the first and fifth metatarsal heads), among eight different conditions of support can be identified. Using a digital-to-analog converter, it is then possible to go from three digital signals to eight analog levels, obtaining a so-called “8-level basography”. However, during a normal gait of a healthy subject, only six of these eight levels can be observed (Figure 1.1a). Moreover, the 8-level basography may result too detailed and variable to be used in practice for gait analysis. Hence, it is simplified in order to obtain the corresponding 4-level basography (Figure 1.1b), in which the four levels correspond to heel contact (H), flat foot contact (F), push off (P) and swing (S), respectively. More specifically:

- during H phase only the foot-switch under the heel is closed;
- during F phase the foot-switch under the heel is closed and at least one of the foot-switches under the forefoot is closed;
- during P phase the foot-switch under the heel is open and at least one of the foot-switches under the forefoot is closed;
- during S phase all the foot-switches are open.

During a walking session for gait analysis, basographic signals of both feet are acquired continuously and they give a “time reference frame” used to evaluate all the other collected signals [8].

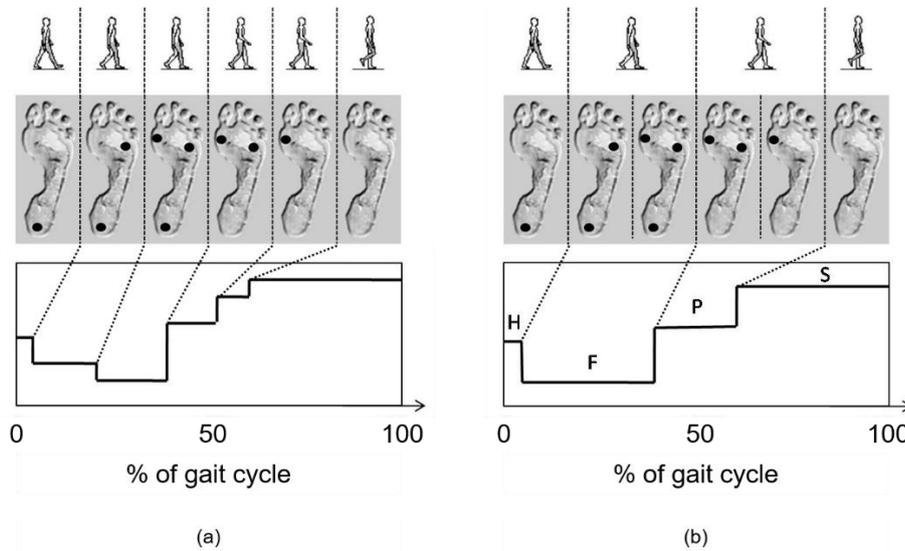


Figure 1.1. Basographic levels.

Representation of (a) six basographic levels derived from an 8-level basography and (b) the corresponding simplified 4-level basography.

1.2.2 Electromyographic signal

The surface electromyography (sEMG) is commonly used for studying the muscle activity during gait [6]. Figure 1.2 shows an example of sEMG acquisition from the Tibialis Anterior muscle of a healthy adult during gait: the signal represents the muscle electrical activity during four consecutive strides (gait cycles).

In sEMG the choice of the electrodes to be used – and the inter-electrode distance – is a critical aspect and it strongly depends on the anatomy of the muscle of interest and its relationship with the other neighboring muscles. Commonly, surface probes consist of two or three detection surfaces of conductor material and, in active probe, an amplifier stage is positioned close to electrodes. At present, active probes are preferred because of their ease of use and because they guarantee better performance with respect to the passive ones.

In general, to obtain high quality EMG recordings, it is a good practice to have different probes and to choose the proper one based on the anatomical characteristics of the muscles of the specific subject. More details about the different EMG probes used in gait analysis can be found in [8].

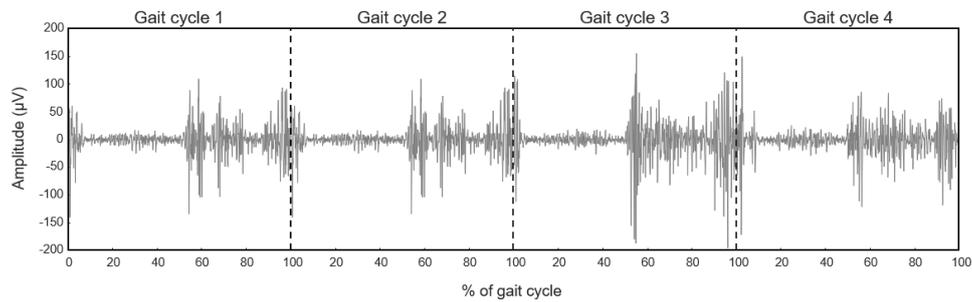


Figure 1.2. Example of EMG signal.

EMG signal of Tibialis Anterior muscle during four consecutive strides, acquired from a healthy adult (male, 51 years old).

1.3 Statistical gait analysis

In Figure 1.3 the activity of Tibialis Anterior muscle of four strides, acquired during a walking session of a healthy adult is reported. Muscular activation intervals are highlighted with a red square. As it emerges from the figure, even considering only a few strides, the muscle activation results quite different from one stride to another: a high variability both in the number of activation intervals and in the duration of activation itself is present. The richness and the variability of muscle activation patterns during gait has been studied and assessed in literature. In [9], sEMG signals of eight muscles of lower limbs were acquired from 25 healthy subjects between the age of 20 and 40 years. All the subjects walked indoors, barefoot, at several different self-selected walking speeds; a minimum of 10 strides were acquired at each speed. The resulting EMG activation patterns demonstrated a considerable amount of inter-individual variability, at each speed. Other evidence of the EMG intra-subject variability are reported in [10], [11] and [12].

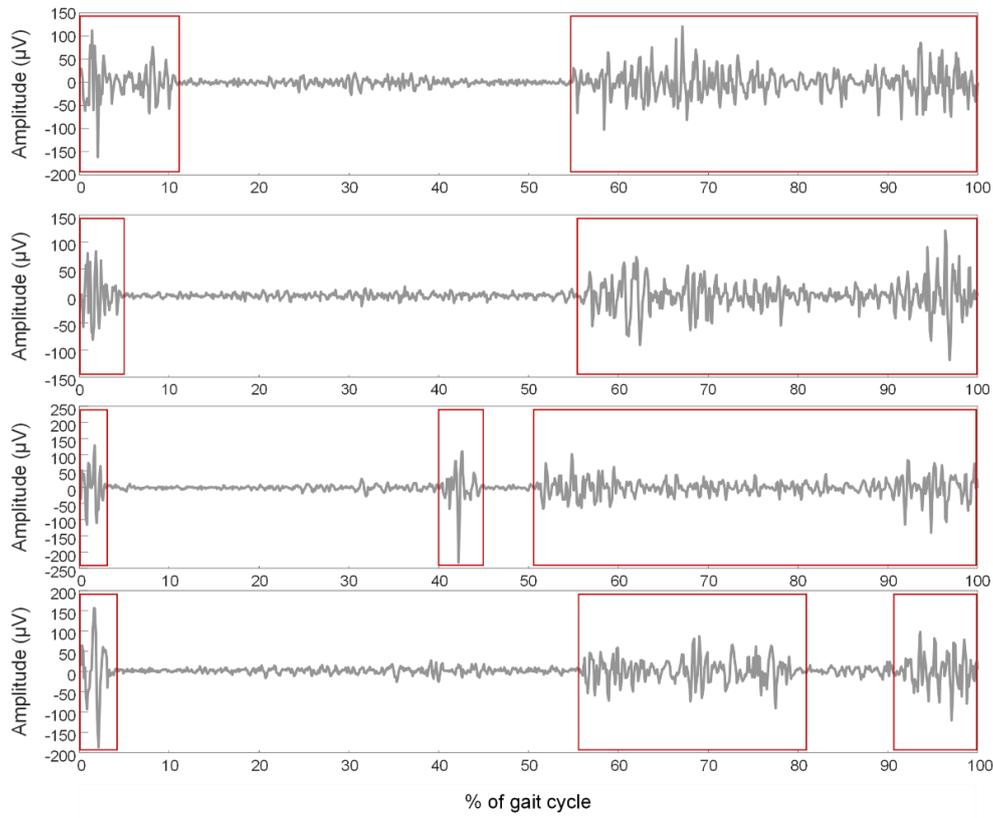


Figure 1.3. Example of intra-subject variability in EMG signal.

Tibialis Anterior EMG signals extracted from a 2-minute walk of a healthy adult (male, 51 years old). Activation intervals are highlighted with red square.

Hence, to deal with this variability, hundreds of strides are needed to be acquired to correctly study gait behaviour. In this context, a powerful methodology that allows for preserving the information during the acquisition of so many consecutive strides is the statistical gait analysis (SGA).

The SGA is based on the “statistical” characterization of gait parameters, both spatio-temporal and EMG derived, and it is used to describe gait functionality in condition similar to the everyday life walking. Systems used in SGA are user-independent and do not hinge on the laboratory characteristics.

Notice that, to guarantee high repeatability of the results, it is convenient to study only those strides related to a straight path: it is necessary to have a walking platform of at least 10 meters that will be walked over a certain number of times [8]. During the acquisition the subject walks back and forth the platform, so it is

then necessary to identify and remove those strides related to acceleration, deceleration and changes of direction that are different from those strides relative to “re-gime” walk.

1.3.1 Segmentation and classification of gait cycle

A crucial point in SGA is the automatic segmentation and classification of gait cycles, since, during a walk, different basographic cycles can be observed. In healthy subjects, the most frequent basographic cycle is represented by the HFPS sequence that can be considered as the typical foot-floor contact sequence. However, a small percentage of “atypical” cycles showing a not-HFPS sequence can be observed (usually less than 5% of the total number of gait cycles). Figure 1.4 shows an example of the basographic cycles obtained during a walking session acquisition of a healthy adult: 144 HFPS cycles, 2 PFPS cycles, 1 FPS and 1 HFP cycle, are classified for left side, while 141 HFPS cycles, 4 PFPS cycles, 1 FPS and 1 HFH cycle, are classified for the right side.

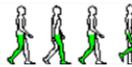
Left side			Right side		
Step ID	Number of strides	Step type	Step ID	Number of strides	Step type
HFPS	144		HFPS	141	
PFPS	2		PFPS	4	
FPS	1		FPS	1	
HFP	1		HFH	1	

Figure 1.4. Example of basographic classification.

Classification of basographic cycles of a healthy subject (male, 51 years old).

SGA can be performed only after the classification of the basographic cycles in the various sequence typologies, since the analysis must be carried out separately for the different type of basographic cycles. Therefore, due to the high number of strides necessary for SGA, it is important to segment and classify gait cycles in a

user-independent and automatic way. In [13] an automatic algorithm for segmentation and classification of gait cycles is presented.

1.3.2 Observation of muscle activations

Several recent studies in literature [14]-[16] have shown that there is a great intra-subject variability in the EMG activation patterns, even in healthy subject performing a single walking session along a straight path. This behaviour is particularly interesting in clinical scenarios, hence it is desirable to obtain, for each muscle, a representation of the different activation patterns characteristic of the subject walking and their frequency. For this purpose, it is necessary to apply proper algorithms for the detection of onset and offset time instants of muscle activity. In the last decades, several approaches have been proposed to solve this challenging task [17]-[20].

Once onset and offset timings have been detected, the identified activation intervals has to be time-normalized with respect of gait cycle duration and then, in order to perform a proper data analysis, it is necessary to group strides showing the same number of activation intervals and consider each group separately.

1.4 STEP32: a movement analysis system

STEP32 (Medical Technology, Italy) is a multichannel wearable system for the acquisition and analysis of gait data. It is able to document in an objective and repeatable way the temporal and space parameters that characterize the gait, such as:

- the sequence and duration of each gait phase, using basographic sensors placed under the foot;
- the neuromuscular activation intervals of the analysed muscles, through the analysis of the raw electromyographic signals acquired with surface electromyographic probes.

Each sensor is cable-connected to a “patient unit”, extremely compact in size (60mm x 30mm x 150mm, weight: 300g) directly applied to the patient by means of an elastic belt. During a walking acquisition session, the data acquired by the patient unit are transferred, through a cable, to a PC and then processed using a dedicated software. The analysis of all the gait parameters is performed using as time reference the sequences and the duration of the basographic cycles.

1.4.1 Sensors

The basographic sensor consists of a rectangular membrane switch (10mm of side), placed at one end of a flexible and insulating plastic strip (Figure 1.5). A connector is placed at the opposite end, necessary for the connection to the preamplifier/decoder. The connector can receive up to three basographic contacts and provides a signal at coded levels in relation to the number of contacts pressed at any time.



Figure 1.5. Basographic sensors connected to preamplifier/decoder.

The probes used for acquiring the electromyographic signals incorporate a differential amplifier stage. They guarantee signal stability, and they reduce movement artifacts, disturbances to the network frequency and the microphonic phenomena of the connection cable. Two types of EMG probes are available (Figure 1.6): fixed geometry single differential probes (the two acquisition electrodes are incorporated in the probe positioned at a fixed distance) and variable geometry probes (the two acquisition electrodes are separated from the pre-amplification part and can be positioned on the patient skin at variable distance).

The electromyographic signal can be amplified from a minimum of 1000 to a maximum of 50.000 times.

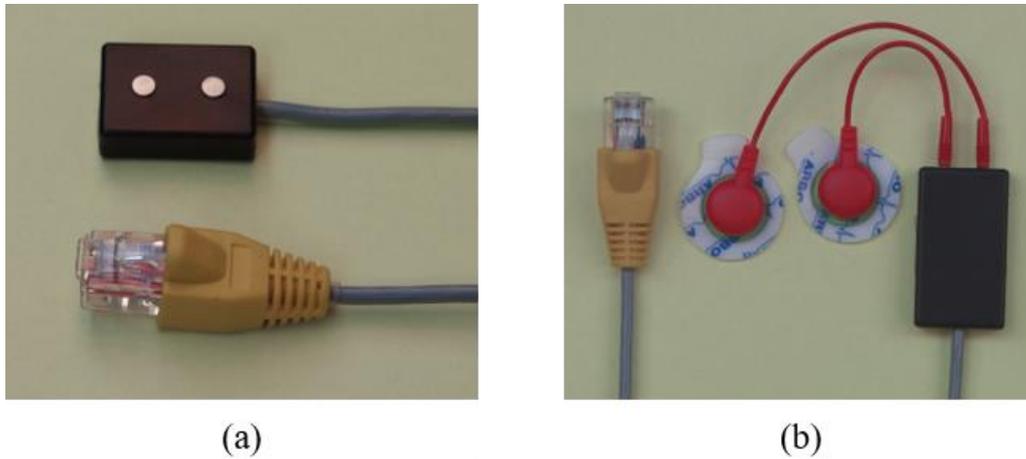


Figure 1.6. EMG probes.

Example of fixed geometry single differential probe (a) and variable geometry probe (b).

1.4.2 Basographic and electromyographic analysis

The routine included in STEP32 for data analysis allows for obtaining the number of cycles showing a specific sequence of gait phases (i.e. Heel contact, Flat foot contact, Push-off, Swing: HFPS cycles) and the average cycle duration for each side. Once selected the desired basographic sequence, the EMG signals are processed using the routine for data analysis integrated in STEP32 system. For each muscle, the processing consists of the following pipeline:

1. The signal is segmented into separate gait cycles and time-normalized to the stride duration [21].
2. Only the strides showing a specific sequence of gait phases (for example HFPS) and related to the straight path [15] are selected.
3. For each stride, the ON/OFF activation intervals are identified using a double threshold statistical detector [22].
4. Strides are grouped together according to their number of activation intervals (modalities). This result is represented using a series of histograms (Figure 1.7): each bar represents the frequency with which the identified modalities occur (expressed in percentage).
5. Only those strides belonging to the most represented modality are selected.

6. ON/OFF timings of these strides are averaged in order to obtain a representative activation pattern (Figure 1.8). The mean value, standard deviation and standard error related to each onset and offset timings of the representative activation patterns are computed and expressed as percentage of gait cycle.

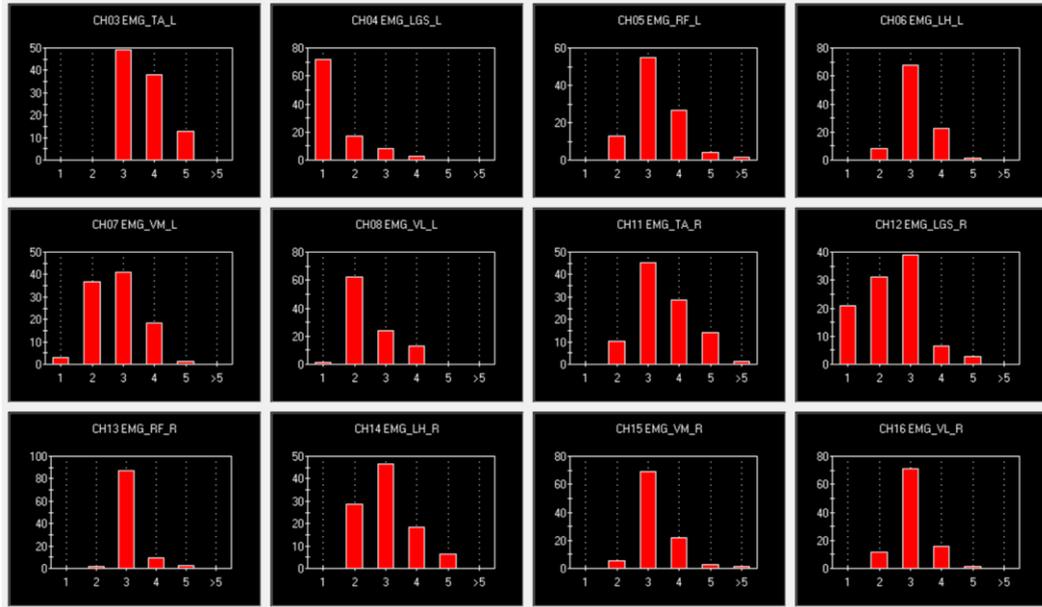


Figure 1.7. Histograms of muscle activity modalities.

Example of histograms representing the frequency (express as percentage) of each modality during a walking session. The analyzed EMG signals were extracted from a walking session of a healthy adult (female, 55 years old). Signals of six muscles were acquired, from both left and right side: Tibialis Anterior (TA), Gastrocnemius Lateralis (LGS), Rectus Femoris (RF), Lateral Hamstring (LH), Vastus Medialis (VM) and Vastus Lateralis (VL).

Hence, using STEP32 system it is possible to obtain several data, both basographic and electromyographic. The aggregation and the correct interpretation of these data is a crucial point to provide an objective and correct assessment of the characteristics of human walking, usable for clinical practice.

In a previous work, a method based on a fuzzy logic classifier, named GAITSCORE, was developed to aggregate data collected from basographic sensors: the result of the system is a gait impairment score, that provide an objective and quantitative measurement of gait impairment [23]. Particularly, 8 basographic parameters for each lower limb were extracted from the foot-switch signal recorded during a subject's walk. These parameters were used as inputs for a Mamdani classifier, based on 36 rules, for the classification of gait. The classifier returned one

output for each lower limb and then they were aggregate into a unique score representing the global walking impairment of the subject.

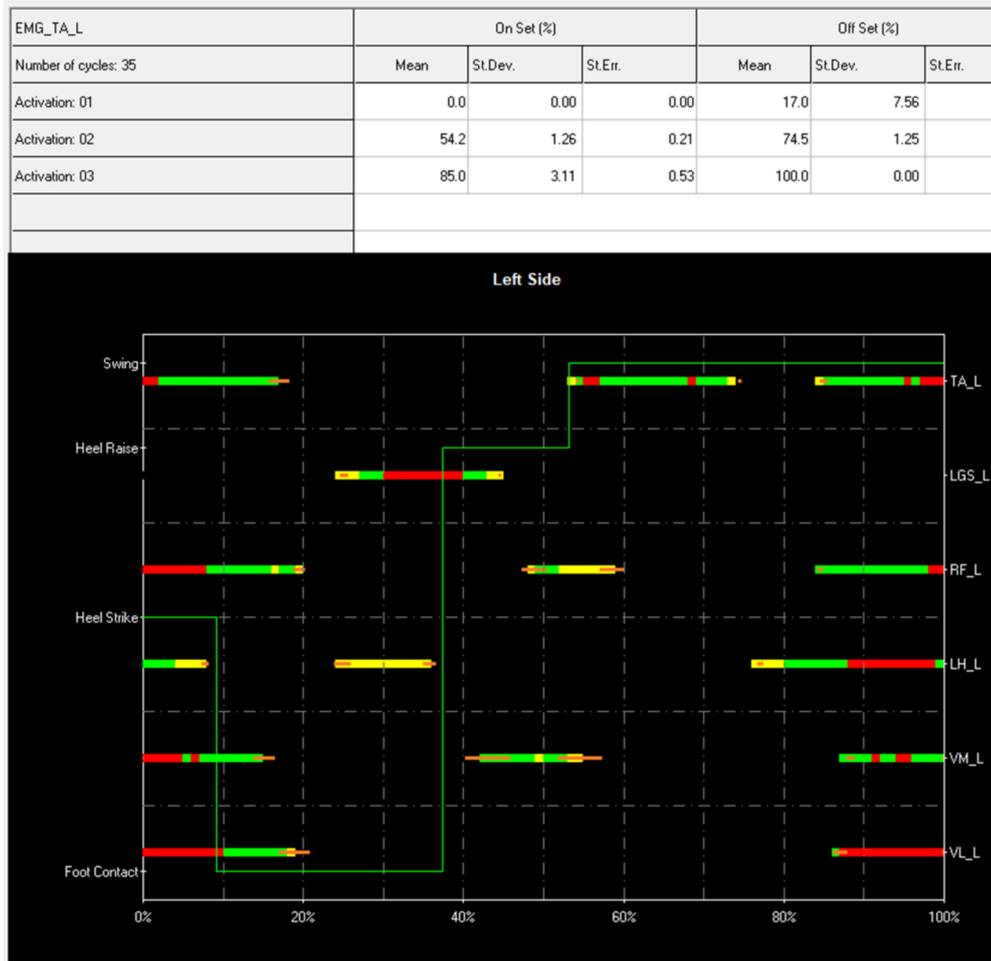


Figure 1.8. Representation of EMG activation intervals (example 1).

Example of representation of EMG activation intervals resulting from the analysis of a walking session of a healthy adult (female, 55 years old). Activation intervals of six muscles (left side) are represented: Vastus Lateralis (VL), Vastus Medialis (VM), Lateral Hamstring (LH), Rectus Femoris (RF), Gastrocnemius Lateralis (LGS) and Tibialis Anterior (TA). In the table are reported, for TA muscle, the mean values, the standard deviation and the standard error of the onset and offset timings, computed considering all the strides presenting a 3-activation modality. In the image, activation intervals are represented with the thicker segment (green, red, yellow), while standard deviation is represented with the orange segment. The different colors with which the activation intervals are represented refers to the different activation intensity (green: low intensity, yellow: medium intensity, red: high intensity).

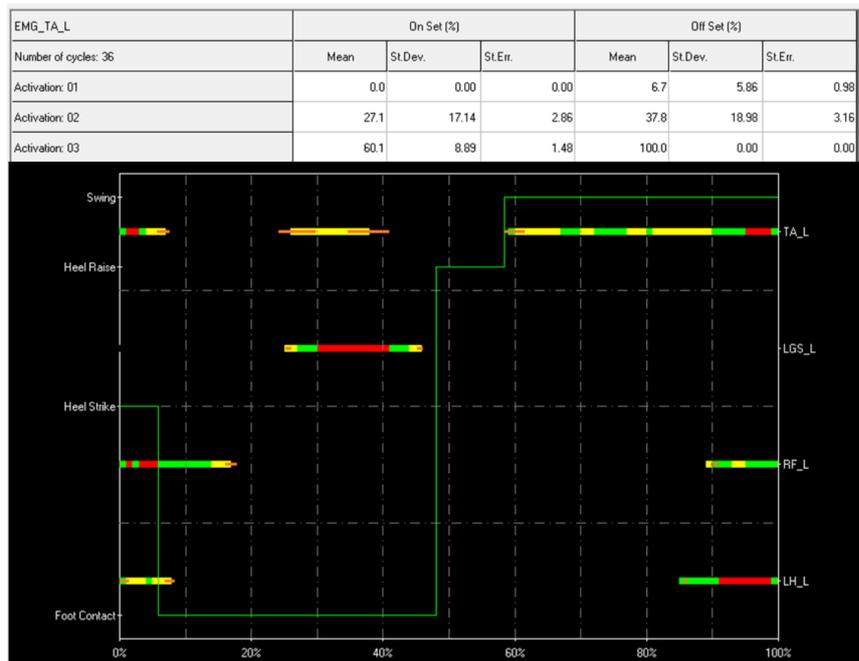
A further improvement of the GAITSCORE system would be the integration of the information related to the timing of muscular activation during gait. However, human locomotion is characterized by a large intra-subject variability in muscle activation patterns, as already discussed and even considering strides showing the standard basographic sequence (HFPS) and belonging to the same activation modality, there are evidences of a great EMG intra-subject variability.

Figure 1.9 shows the EMG results obtained acquiring gait data from a healthy adult (male, 51 years old). Left (Figure 1.9a) and right (Figure 1.9b) activation patterns of four muscles are reported: Lateral Hamstring (LH), Rectus Femoris (RF), Gastrocnemius Lateralis (LGS) and Tibialis Anterior (TA).

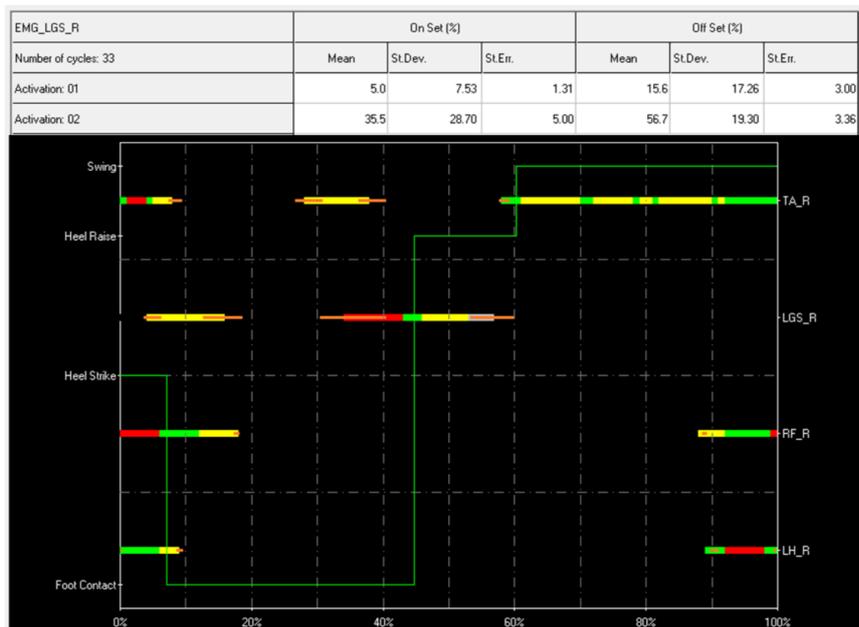
From Figure 1.9a emerges that the TA muscle (left side) presents, in the second activation interval, high values of standard deviation related to both the onset and offset timings (17.14 and 18.98% of gait cycle, respectively).

Similarly, from Figure 1.9b, emerges that the LGS muscle (right side) presents, in the first activation interval, a high value of standard deviation related the onset timing (28.7% of gait cycle). Moreover, in the second activation interval, high values of standard deviation are found related both to the onset and offset timing (17.26 and 19.3% of gait cycle, respectively).

Hence, in addition to the variability in the number of modalities that a subject expresses during walking [14], [24], it has to take into account that also different types of activation pattern are present, even considering the same modality (Figure 1.10). Therefore, for a correct interpretation of EMG signals acquired during walking and a correct aggregation of EMG data into GAITSCORE system, it is necessary to adopt proper strategies for grouping cycles sharing not only the same activation modality, but also similar activation patterns.



(a)



(b)

Figure 1.9. Representation of EMG activation intervals (example 2).

Examples of EMG activation intervals resulting from the analysis of a walking session of a healthy adult (male, 51 years old). Activation intervals of four muscles are represented: Lateral Hamstring (LH), Rectus Femoris (RF), Gastrocnemius Lateralis (LGS) and Tibialis Anterior (TA). The tables report, for (a) TA muscle (left side) and (b) LGS muscle (right side), the mean value, the standard deviation and the standard error of the onset and offset timings of each activation interval.

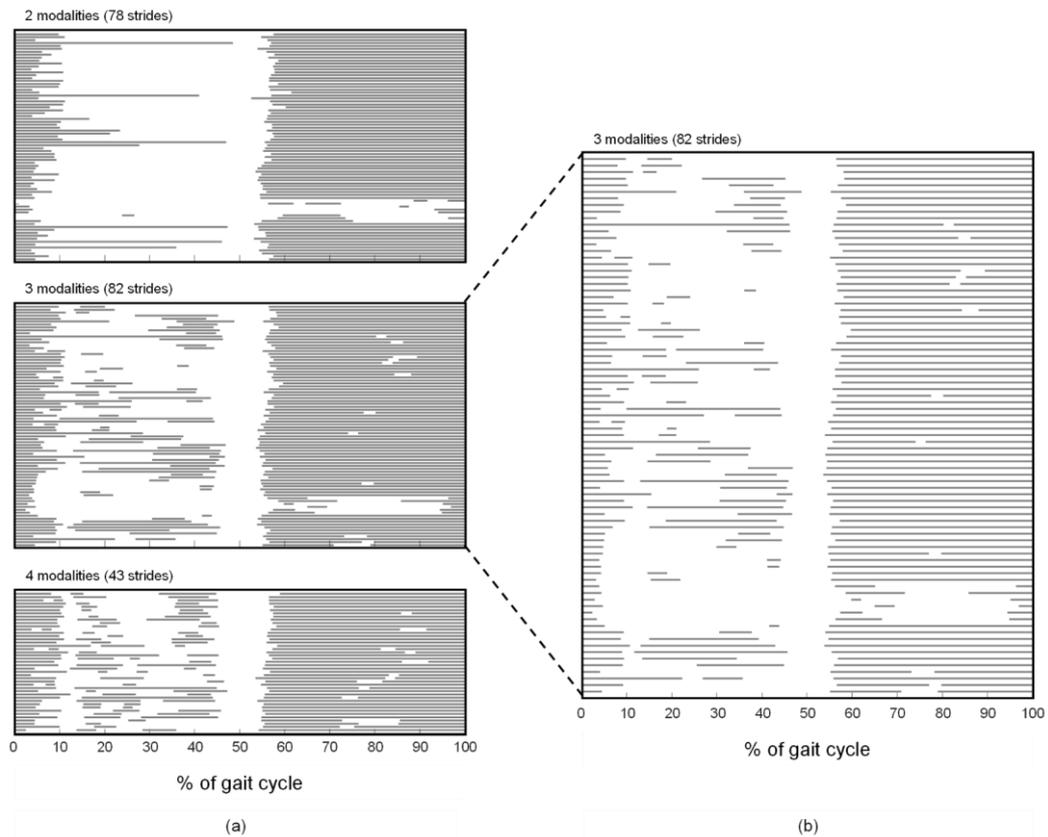


Figure 1.10. Example of EMG pattern variability.

a) Strides are grouped according to their number of activation intervals (modalities). b) Strides belonging to the modality with 3 activation intervals are grouped together: a great intra-modality variability in the EMG activation patterns is evident.

Chapter 2

CIMAP: Clustering for Identification of Muscle Activation Patterns

2.1 Introduction

In order to cope with the issues presented in the previous chapter, a method to group gait cycles showing homogeneous EMG activation patterns was developed.

Clustering methods are a class of methods suitable for this purpose, because they are able to identify groups of elements among a dataset, sharing similar characteristics. The similarity among the elements is evaluated using a distance metric. The goodness of the clustering result is evaluated assessing the variability among the elements assigned to the same cluster (it should be as low as possible), and the distance between the resulting clusters (the more the distance between clusters is large, the more the clusters are clearly distinguishable).

Among the existing clustering methods, the most suitable solution for the problem is the *dendrogram*, an agglomerative hierarchical clustering method that organizes data in a hierarchical tree based on a proximity measure [25]. Particularly, the agglomerative hierarchical clustering is a method that works using a "bottom-up" approach: each element of the dataset starts in its own cluster, and pairs of clusters are merged at each iteration. The hierarchical relationship between elements is most commonly represented using a dendrogram diagram (cluster tree). The elements are allocated to final clusters by "cutting the tree" at a certain level according to specific defined criteria.

In literature, there are a few works in which clustering methods have been applied to EMG data analysis and they are limited to three main fields: segmentation [26] and decomposition [27] of EMG signals, analysis of subjects' populations [28], and the analysis of subject variability during different walking sessions [29]. The clustering approach has never been applied to deal with the intra-subject variability in the EMG activation patterns.

The CIMAP (Clustering for Identification of Muscle Activation Patterns) is the method, based on agglomerative hierarchical clustering, developed for grouping strides with similar EMG activation patterns [30], [31].

In this chapter and in the following sections of the thesis, several datasets are used to test and validate the proposed algorithms. In Appendix A, a full description of these datasets can be found.

2.2 Method

Figure 2.1 shows the CIMAP processing pipeline. The pipeline of the method is based on two main steps:

1. EMG signal pre-processing, consisting of signal segmentation, stride extraction and activation interval detection.
2. Implementation of the CIMAP algorithm, consisting of dataset preparation and clustering of each dataset.

2.2.1 EMG signal pre-processing

The EMG signal pre-processing step consists of:

1. *Signal segmentation.* EMG signals collected during a walking session are segmented into separated strides and time-normalized to the stride duration [13].
2. *Stride extraction.* Among all the strides, only those corresponding to a specific foot-floor contact type are extracted (i.e. HFPS strides). Moreover, those strides related to deceleration, turning, and acceleration in correspondence of direction changes are discarded using a multivariate statistical filter [15].
3. *Activation intervals detection.* The double threshold statistical detector described in [22] is used to identify the activation intervals of each stride previously extracted.

As a result of the pre-processing step, a collection of separated strides is obtained and, for each stride, the onset-offset timings of EMG activity are detected.

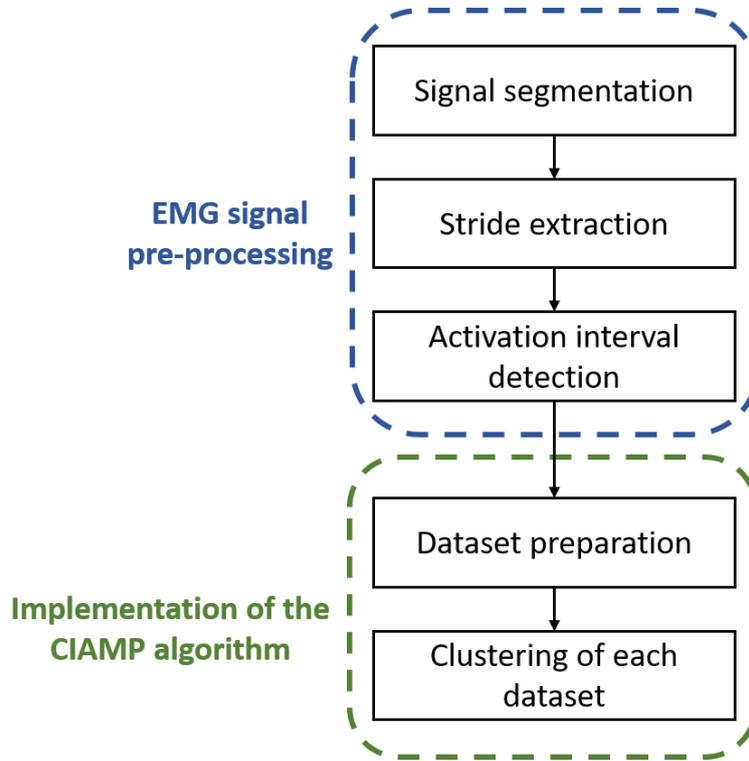


Figure 2.1. CIMAP processing pipeline.

2.2.2 Implementation of the CIMAP algorithm

The CIMAP algorithm is applied separately on each muscle, following two phases: a) datasets preparation, and b) clustering of each dataset.

Dataset preparation

The strides of both lower limbs are joined to obtain a higher number of gait cycles for the clustering procedure. Strides are then split into different datasets according to their activation modality. More specifically, every stride i of a specific dataset is identified by m couples of time instants (ON_j, OFF_j) where m is the number of activation intervals within the stride:

$$stride_i = \{ON_{i,1}, OFF_{i,1}, \dots, ON_{i,j}, OFF_{i,j}, \dots, ON_{i,m}, OFF_{i,m}\} \quad (2.1)$$

For the further analysis, only datasets consisting of at least three strides are considered.

Clustering of each dataset

The clustering procedure is applied separately to each dataset. At first, each stride represents a single-element cluster. Then, the cluster tree is constructed by merging, at each iteration, the two closest clusters. Particularly, each cluster (A) is characterized by its centroid (CLC_A) defined as a vector containing the mean values of the ON and OFF timings of all the strides belonging to the cluster:

$$CLC_A = \frac{1}{|A|} \sum_{a_i \in A} a_i = \{\overline{ON_1}, \overline{OFF_1}, \dots, \overline{ON_j}, \overline{OFF_j}, \dots, \overline{ON_m}, \overline{OFF_m}\} \quad (2.2)$$

In (2.2), $|A|$ is the cardinality (number of strides) of the cluster A and a_i is the i -th element (stride) belonging to cluster A .

At each iteration, the two clusters with the closest centroids are merged. The *Chebyshev distance* is used as distance metric to assess the similarity between two clusters A and B :

$$d(CLC_A, CLC_B) = \max\left(\max_m(ON_A - ON_B), \max_m(OFF_A - OFF_B)\right) \quad (2.3)$$

The *Chebyshev distance* was chosen to avoid merging elements showing a single, markedly different, onset (or offset).

Finally, an automatic rule consisting of three steps is used to identify the dendrogram cut point:

- 1) At each iteration k , when two clusters (A and B) are merged, the R_k index is computed as follows:

$$R_k = \frac{\sum_{stride_i \in (A \cup B)} dist(stride_i, CLC_{A \cup B})}{\max(\sum_{stride_i \in A} dist(stride_i, CLC_A), \sum_{stride_j \in B} dist(stride_j, CLC_B))} \quad (2.4)$$

The index represents the ratio between the intra-cluster variability of the new formed cluster (AUB) and the maximum between the intra-cluster variability of the two original clusters A and B . The intra-cluster variability is computed as the sum of all the Euclidean distances between each element (stride) belonging to the cluster and the cluster's centroid. The R_k values are low at the beginning of the merging process and grows gradually in the subsequent iterations.

- 2) The R_k values only of the last 20% iterations are considered, to avoid cutting the tree at a very low level.
- 3) The iteration k with the maximum R_k value is identified and the tree is cut at iteration $k-1$, obtaining the clustering result.

For each cluster, left and right strides are divided, obtaining two different clusters (one per side). Finally, clusters with less than 5% of the total number of strides (of the side under consideration) are discarded.

2.3 Testing

A dataset consisting of twenty healthy subjects aged from 20 to 51 years (7 males/13 females, height: 167 ± 8 cm; weight: 62.5 ± 10.7 kg) was collected and used as *Test Set* for testing the CIMAP method. All the subjects walked barefooted at self-selected speed back and forth along a 10-m walkway for at least 2.5 min.

The acquisition system STEP32 for gait analysis was used to acquire basographic signals and surface EMG signals. Three foot-switches were placed under the foot-soles, beneath the first and fifth metatarsal heads, and beneath the back portion of the heel. After skin preparation, the surface EMG probes were placed bilaterally over the muscle's belly of the analyzed muscles. EMG probes were active and utilized AgCl-disks as electrodes. The signal amplifier had a gain of 1000 and a 3-dB bandwidth from 10 Hz to 400 Hz. The sampling frequency was 2 kHz and signals were converted by a 12-bit analog to digital converter.

Since the purpose of this first phase of the study was to verify that the CIMAP algorithm could help with the problem of EMG intra-subject variability, for each

subject only one muscle (Tibialis Anterior - TA) was analyzed and only the modality with the higher number of strides, resulted using STEP32 analysis tool, was considered.

To compare the stride variability with and without clustering, the average intra-cluster variability among strides belonging to the same cluster was compared with the average variability among strides belonging to the same activation modality. Particularly, for a specific cluster, the intra-cluster variability was defined as the average standard deviation (SD) of the ON and OFF timings among the cluster's strides:

$$Cluster_SD = \left[\overline{SD(ON_1), SD(OFF_1), \dots, SD(ON_m), SD(OFF_m)} \right] \quad (2.5)$$

For each subject's lower limb, the average *Cluster_SD* was calculated across all the resulting clusters and then, *Cluster_SD* values of left and right side were also averaged.

The variability among strides sharing the same activation modality was calculated in the same manner. The intra-modality variability was defined as the average SD of the ON and OFF timings among the modality's strides:

$$Modal_SD = \left[\overline{SD(ON_1), SD(OFF_1), \dots, SD(ON_m), SD(OFF_m)} \right] \quad (2.6)$$

For each subject's lower limb, the *Modal_SD* was calculated across the selected modality, and then, the values of both sides were also averaged.

In Figure 2.2 an example of clustering result is reported, both for left and right side. As it emerges from the figure, the CIMAP algorithm allows for extracting the representative activation patterns of a subject's muscle, each corresponding to a cluster centroid. Even considering only the most represented activation modality, several clusters were obtained, confirming the richness of muscle activation patterns during gait, already documented in literature [32], [33], [34].

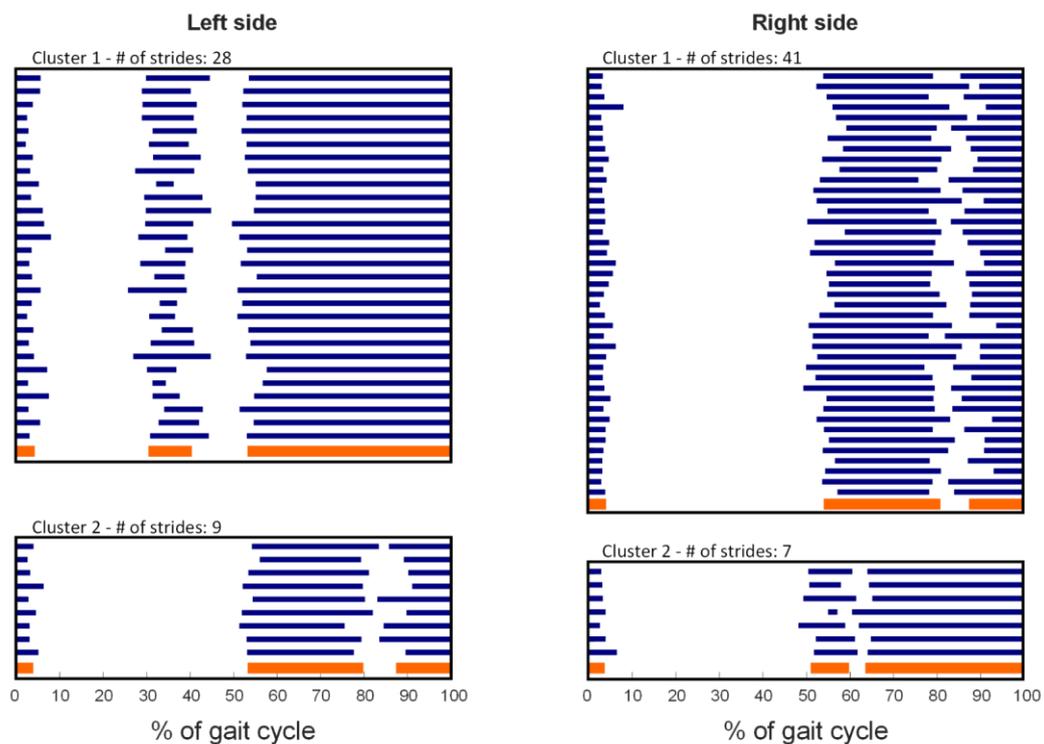


Figure 2.2. Example of clustering result.

Example of clusters obtained from a healthy adult using the CIMAP algorithm. Muscles: Tibialis Anterior (both sides). In each row, the blue bar represents the EMG activation intervals within a gait cycle (BLUE = muscle active, WHITE = muscle inactive). The number of strides of each cluster is reported above the corresponding plot area. Below each cluster, the cluster centroid is represented in orange.

Figure 2.3 shows the boxplots of the variability values ($Cluster_SD$ and $Modal_SD$) computed before and after the clustering procedure. The mean value and standard deviation of the variability values obtained in the two cases, are equal to 2.0 ± 0.6 (with clustering) and 8.1 ± 1.8 (without clustering). A paired Wilcoxon test ($\alpha=0.05$) was then applied, and it proved that the variability is significantly reduced after clustering ($p < 0.001$).

These results proved that the CIMAP method could be a promising tool for helping in the interpretation of EMG data in gait analysis, by grouping strides with similar activation patterns.

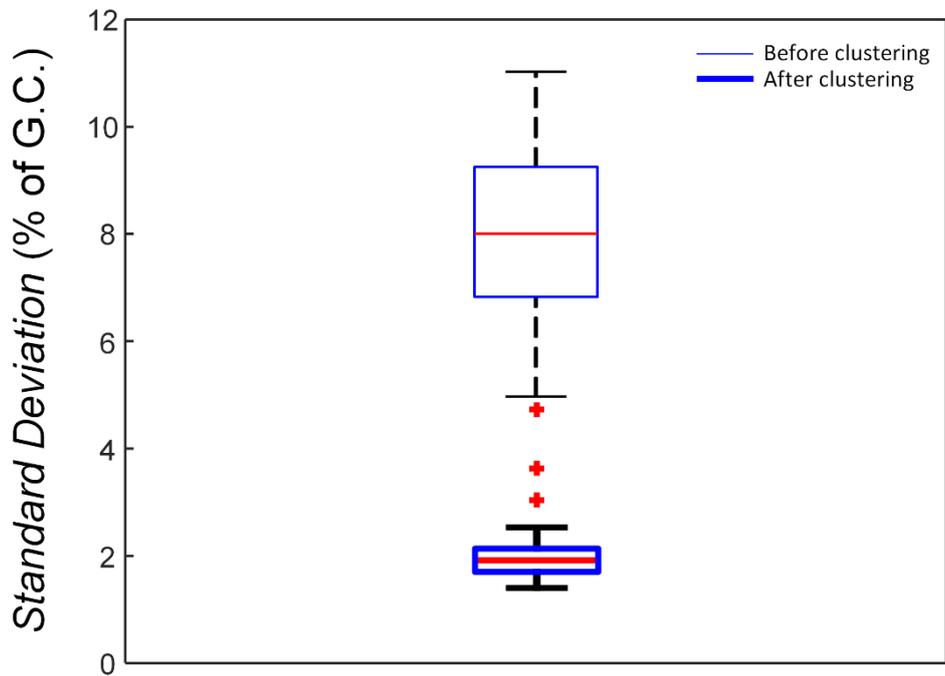


Figure 2.3. Mean cluster SD.

Comparison of the variability in the activation onset/offset among strides grouped in (a) activation modalities (before clustering) or (b) clusters (after clustering), for Tibialis Anterior (TA) muscle. Each boxplot represents the value distribution across the sample of 20 healthy subjects.

2.4 Validation

To validate the CIMAP algorithm, the study was extended to the analysis of patients suffering from different locomotion pathologies (both orthopedic and neurological). The aim of this phase was to test CIMAP performances in critical conditions, since, in these patients, was expected a higher EMG intra-subject variability with respect to healthy subject.

Gait data of 20 patients aged from 16 to 85 years (10 males/ 10 females, height: 168.0 ± 10.7 cm; weight: 73.5 ± 13.9 kg) suffering from four different locomotion alterations were analyzed. In particular, the *Validation Set* consisted of:

- 5 patients after total knee replacement (TKR);
- 5 patients with megaprosthesis of the knee after bone tumor resection (Mega TKR);
- 5 patients with total hip arthroplasty (THA);
- 5 patients affected by normal pressure hydrocephalus (iNPH).

TKR, Mega TKR and THA are the consequence of orthopedic diseases, while iNPH is a neurological disease.

All subjects walked barefoot, at self-selected speed, consecutively for 2-3 minutes. The wearable system STEP32 was used to acquire gait data, with the same acquisition protocol previously detailed.

The variability with and without clustering among the new dataset was assessed using *Cluster_SD* and *Modal_SD*. Figure 2.4 shows the boxplot of the variability before and after clustering. The mean value and standard deviation of the variability values are equal to 3.1 ± 2.1 (without clustering) and 9.3 ± 4.6 (with clustering).

The paired Wilcoxon test ($\alpha=0.05$) was applied, and the variability resulted significantly reduced after clustering ($p < 0.001$). However, it is evident a less marked difference between the two distributions, with respect to the results obtained on healthy subjects.

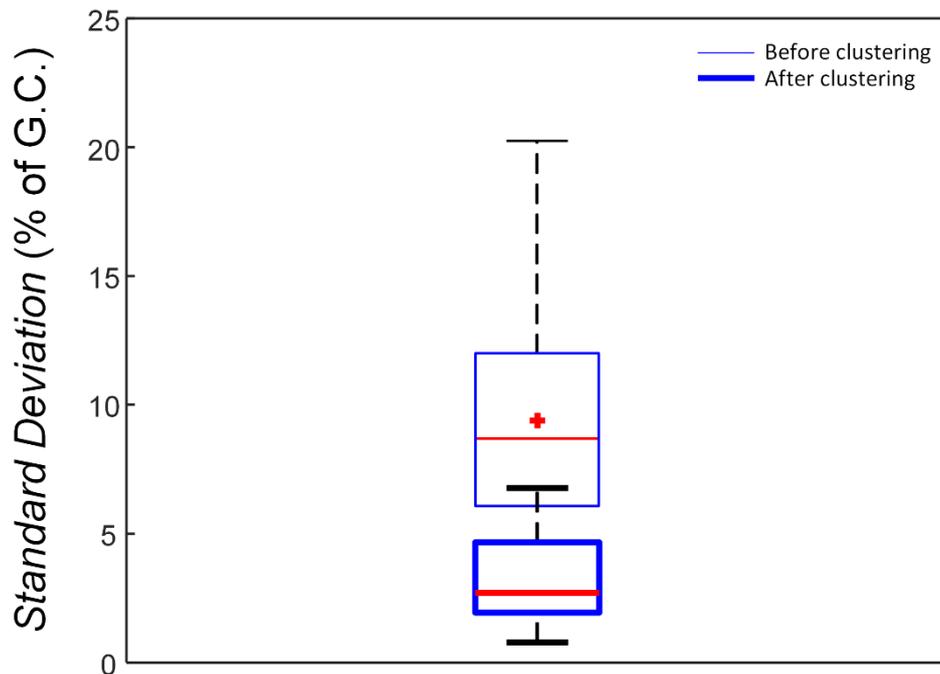


Figure 2.4. Mean cluster SD.

Comparison of the variability in the activation onset/offset among strides grouped in (a) activation modalities (before clustering) or (b) clusters (after clustering), for Tibialis Anterior (TA) muscle. Each boxplot represents the value distribution across the sample of 20 patients.

Figure 2.5 shows the result of the application of the CIMAP algorithm to an iNPH patient. As it emerges from the figure, the CIMAP algorithm allows for extracting different activation patterns as well as previously reported, but a great intra-cluster variability is present in this condition.

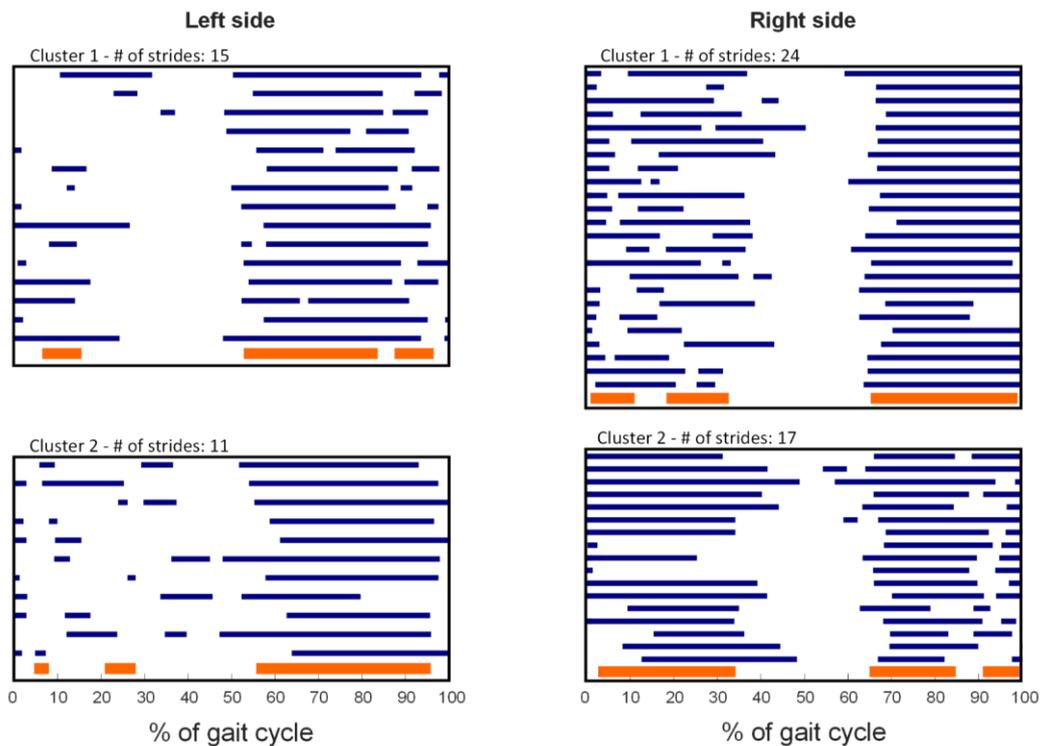


Figure 2.5. Example of clustering result.

Example of clusters obtained from an iNPH patient. Muscles: Tibialis Anterior (both sides). In each row, the blue bar represents the EMG activation intervals within a gait cycle (BLUE = muscle active, WHITE = muscle inactive). The number of strides of each cluster is reported above the corresponding plot area. Below each cluster, the cluster prototype is represented in orange.

The results obtained on the *Validation Set* highlighted some issue to be solved in the CIMAP algorithm, probably due to a choice of dendrogram parameter suitable in case of low EMG variability (associated to healthy subjects), but not optimal in case of higher EMG variability (associated to patient with locomotion alterations).

Chapter 3

Optimization of the CIMAP algorithm

3.1 Introduction

In the previous chapter, the CIMAP algorithm was introduced to overcome the limitation of gait data interpretation due to the high variability of the EMG activation patterns. The proposed algorithm was developed and tested on a set of EMG data acquired from healthy subjects. The results obtained on these subjects showed that the CIMAP was able to reduce the cycle-to-cycle variability, providing more representative EMG patterns.

However, it was observed that the algorithm performances decreased when it was applied to gait data acquired from pathological subjects (*Validation Set*), producing clusters with high intra-cluster variability.

In this chapter, the optimization of the CIMAP algorithm is presented, in order to obtain clusters consisting of gait cycles with similar activation patterns, both in healthy and pathological subjects [35].

3.2 Tuning of the dendrogram parameters

In order to identify the best configuration for the CIMAP algorithm implementation (in terms of centroid definition, linkage method and selection of the cutoff

point), several dendrogram settings were tested and their influence on the final algorithm performance were evaluated.

3.2.1 Centroid definition

The *Centroid* is defined as the element that characterizes each cluster. In the CIMAP method, the centroid was defined as a $2m$ -dimensional vector containing the “representative” ON and OFF timings (where m represents the modality).

In order to optimize the method, two different centroid definitions were tested: the mean value (CLC_{mean}) and the median value (CLC_{median}) of the elements belonging to the cluster.

3.2.2 Linkage method

The *Linkage Method* is the procedure for the selection of the clusters to join at each iteration. Two different linkage methods were tested: the *centroid linkage* (*Centr_Link*) and the *complete linkage* (*Comp_Link*).

The *centroid linkage* merges two clusters by evaluating the distance between their centroids: at each iteration, the distance between every couple of centroids is computed and the two closest clusters are merged together (Figure 3.1a).

The *complete linkage* merges two clusters by evaluating the distance between every couple of elements in the two considered clusters [36]: the distance between every pair of elements in the clusters is calculated and the largest one is considered as the distance between the two clusters. Then, the two clusters with the smallest distance are joined (Figure 3.1b).

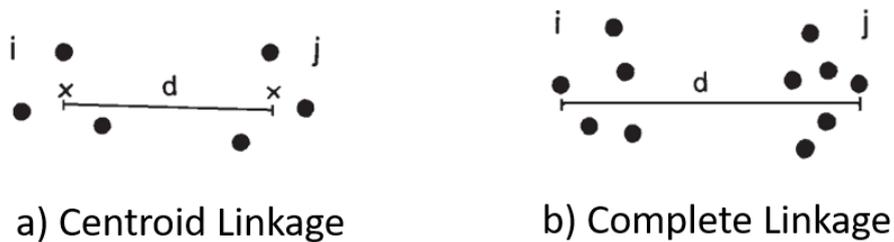


Figure 3.1. Linkage Methods.

Example of Centroid Linkage (a) and Complete Linkage (b) between two clusters (i and j).

3.2.3 Distance metric

The *Distance Metric* is used for assessing the element similarity during the linkage process. A preliminary analysis was performed in order to identify the most promising metrics. After empirical proofs (Appendix B) two metrics were selected: the Manhattan (or Cityblock) distance (D_{man}) and the Chebyshev distance (D_{cheb}).

3.2.4 Selection of the cutoff point

Several tests (Appendix C) were performed to assess the validity of the cutoff rule defined in Chapter 2 for cutting the tree. These tests point out some issues in the cutoff rule that turned out to be not suitable for all the situations. Therefore, it was necessary to introduce a new cutting rule, which would be more suitable to manage the different situations that may arise during the clustering procedure.

In [37], Demirmen proposed some general criteria for the selection of groups in clustering problems:

1. Any chosen group must stand out from neighboring classes: distance between groups should be maximized.
2. Whereas an individual group should not contain a large proportion of the available samples, it is possible to produce a number of groups containing only a small number of samples.
3. The number of groups evolved must be of practical use.

Starting from these considerations, a new cutoff point based on a rule that consists of three criteria is defined. Particularly, the criteria are based on the distance between the merged clusters (*Diff*): considering the series of differences (*Diff*) between consecutive iteration, the three cutoff points were defined as follows:

- *CutA*: first iteration in which the difference *Diff* is higher than the average difference μ_{Diff} (Fig. 3.2a).

- *CutB*: first iteration in which the difference *Diff* is higher than $\mu_{Diff} + \sigma_{Diff}$, where σ_{Diff} is the standard deviation of *Diff* (Fig. 3.2b).

- *CutC*: a moving average (window: 5 points) is applied to the *Diff* series. Starting from the last value and stepping backwards, the cutoff is identified as the point in which the series stop decreasing monotonically (Fig. 3.2c).

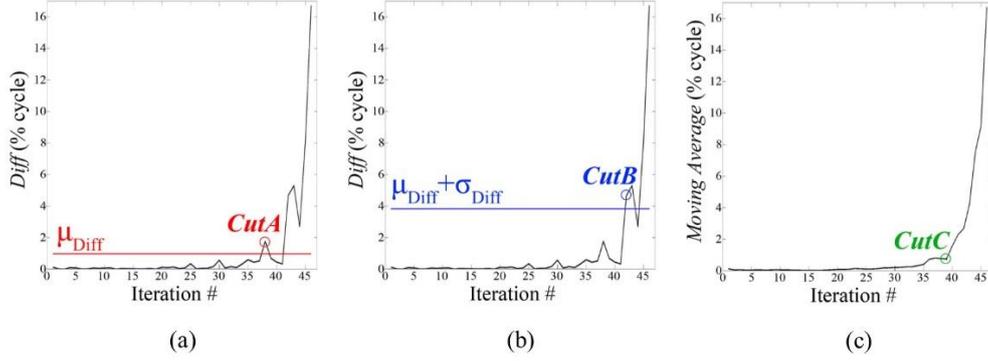


Figure 3.2. Cutoff points.

Representation of the three cutting criteria based on the *Diff* series.

As it emerges from Figure 3.3, the three cutoff criteria may result in a different number of clusters. The best cutoff was identified as the one that is a good compromise between low intra-cluster variability and a high number of cycles included in *significant* clusters (clusters containing at least 10% of the total number of cycles). To automatically select the best cutoff, the following index was defined:

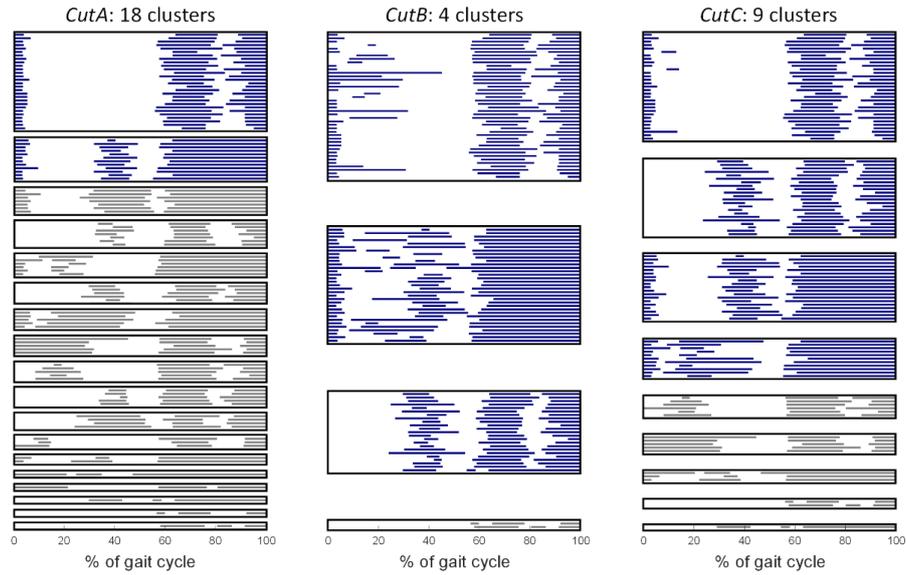
$$CUT_IND = \frac{\sum_{i=1}^n INTRA_VAR_i / n}{\sum_{i=1}^n |C_i|} \quad (3.1)$$

where n is the number of *significant* clusters, $|C_i|$ represents the number of cycles included in cluster C_i , and $INTRA_VAR_i$ is the intra-cluster variability of the i -th cluster calculated as:

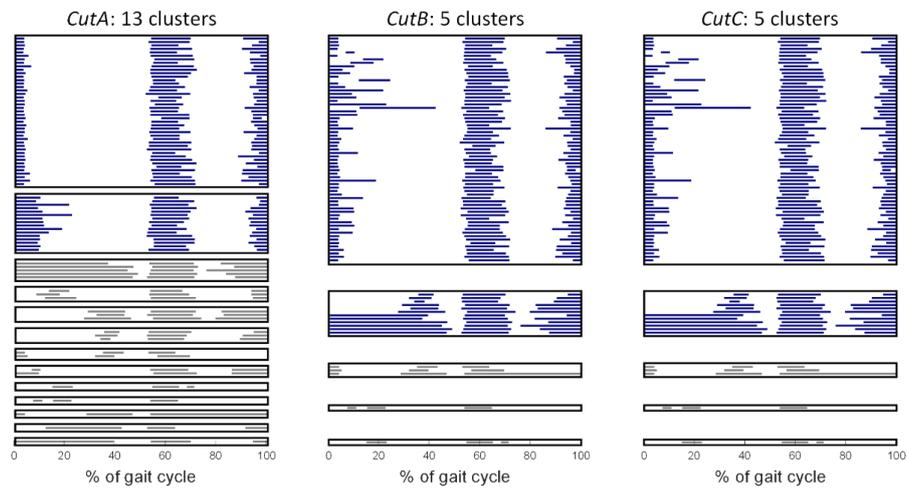
$$INTRA_VAR_i = \overline{dist(cycle_j, cycle_k)}, \forall j, k \in C_i \quad (3.2)$$

where $dist$ is the Cityblock distance.

In case of small intra-cluster variability and high number of cycles included in the significant clusters, the index assumes low values. Therefore, for each dendrogram the three cutoff points are applied and the one that returned the lowest value of the *CUT_IND* index is selected.



(a)



(b)

Figure 3.3. Comparison of three cut-off criteria.

Two examples of the clusters resulting applying the three cut-off criteria defined. Significant clusters are represented in blue; not-significant clusters are represented in grey. In case (a) the criteria provide three different clustering results, while in case (b) *CutB* and *CutC* provide the same clustering result.

3.3 Clustering evaluation

Eight combinations (3 parameters with 2 possible settings each) were tested. For each combination the CIMAP method was applied and, once the dendrogram was cut, the resulting clusters were split into left and right clusters.

The clustering results of each combination were evaluated both for left and right side, considering only the representative clusters (those clusters containing at least the 5% of the total number of strides of the side under consideration). The evaluation was performed by means of a numerical index (*CLUSTER_VAR* index) that evaluate two aspects: the similarity of the strides grouped in the same cluster and the number of strides included in representative clusters. More specifically, for a specific cluster i , the *CLUSTER_VAR* index is calculated by using (3.3):

$$CLUSTER_VAR_i = \frac{\sum_{j=1}^p dist(stride_j, CLC_i)}{p}, \forall j \in C_i \quad (3.3)$$

where p is the number of strides included in the representative cluster C_i , CLC_i is the cluster centroid and $dist$ is the Cityblock distance. The final value of *CLUSTER_VAR* is obtained by averaging the values of all the clusters.

3.4 Testing

The dataset (*Dataset_opt*) consisting of the healthy and pathological subjects used in previous chapter (*Test Set* and *Validation Set*) was used to evaluate CIMAP results using the 8 combinations of parameters, in order to find the best set of parameters. The analysis was extended to all the modalities of each subjects, to increase the number of available datasets.

For each subject, the strides were separated according to their number of activation intervals, obtaining several datasets for each subject (one dataset per modality with more than 2 strides).

Each dataset was used as input for the CIMAP algorithm, using the 8 combinations of parameters. Among all the datasets, only those producing at least one significant cluster for all the combinations were selected: a total of 551 valid datasets were then obtained. For each combination, the number of occurrences (*#OCC*) in which that specific combination outperforms the others was computed, i.e. the final clustering result shows the lowest *CLUSTER_VAR* values.

In Table 3.1 the number of occurrences (#OCC) of each combination is reported; the sum across all the configurations (657) is higher than the total number of analyzed datasets (551) because in several situation two configurations of parameters returned exactly the same results and so both of them were assigned as having the lowest value of *CLUST_VAR*.

Table 3.1. #OCC results (8 combination of parameters)

<i>Centroid</i>	<i>CLC_{mean}</i>				<i>CLC_{median}</i>			
	<i>Linkage method</i>		<i>Linkage method</i>		<i>Linkage method</i>		<i>Linkage method</i>	
	<i>Compl_Link</i>	<i>Centr_Link</i>	<i>Compl_Link</i>	<i>Centr_Link</i>	<i>Compl_Link</i>	<i>Centr_Link</i>	<i>Compl_Link</i>	<i>Centr_Link</i>
<i>Distance metric</i>	<i>D_{man}</i>	<i>D_{cheb}</i>	<i>D_{man}</i>	<i>D_{cheb}</i>	<i>D_{man}</i>	<i>D_{cheb}</i>	<i>D_{man}</i>	<i>D_{cheb}</i>
#OCC	6	5	39	46	177	181	103	100

At first, the centroid type was analyzed: comparing the sum of #OCC obtained by the 4 combinations that use the mean value as centroid (#OCC equal to 96), to those obtained by the 4 combinations that use the median value (#OCC equal to 561), it is evident that *CLC_{median}* outperform the other centroid type.

Then, considering only the combination that used *CLC_{median}*, a preponderance of *Comp_Link* was observed, which obtains about the double of #OCC than *Centr_Link*. This preponderance is also evident when the #OCC are re-computed considering only the four combination that use *CLC_{median}* (Table 3.2).

Table 3.2. #OCC results (4 combination of parameters)

<i>Centroid</i>	<i>CLC_{median}</i>			
	<i>Linkage method</i>		<i>Linkage method</i>	
	<i>Compl_Link</i>	<i>Centr_Link</i>	<i>Compl_Link</i>	<i>Centr_Link</i>
<i>Distance metric</i>	<i>D_{man}</i>	<i>D_{cheb}</i>	<i>D_{man}</i>	<i>D_{cheb}</i>
#OCC	202	208	116	115

Finally, it was observed that the effect of the two distance metrics does not produce any relevant difference in #OCC.

For these reasons, the better configuration of parameters resulted to be the combination of *CLC_{median}* and *Comp_Link*. On the other hand, since no relevant differ-

ence was observed in clustering results using the Manhattan or Chebychev distances, the dendrograms using both these metrics are constructed and the clusters with the lowest *CLUST_VAR* are considered as definitive result.

The two version of the CIMAP (old and optimized one) were compared using the average *Cluster_SD* calculated across all the resulting clusters for each dataset. The *Cluster_SD* values and the number of cycles not included in the significant clusters (*CYC_NSC*) for the two versions are shown in Tables 3.3 (mean \pm standard deviation). A paired Student t-test ($\alpha=0.05$) was performed on each group and the test results proved that the new CIMAP version allows for obtaining cluster with a lower variability with respect to the clusters obtained using the original version ($p<0.001$), for all the groups. Moreover, significant differences ($p<0.001$) were obtained comparing the *CYC_NSC* values between the two versions. The new version of the algorithm produces a reduction in the cluster variability, but a slightly higher number of cycles is discarded (on average, not more than 5 cycles).

Table 3.3. Comparison of Original and new CIMAP

	Cluster_SD (% of g.c.)		CYC_NSC (# cycles)	
	Original CIMAP	New CIMAP	Original CIMAP	New CIMAP
<i>Healthy subject</i>	1.9 \pm 1.3*	1.4 \pm 0.7*	8 \pm 8*	12 \pm 7*
<i>THA patient</i>	2.6 \pm 1.6*	1.5 \pm 0.7*	3 \pm 3*	6 \pm 4*
<i>TKR patient</i>	2.1 \pm 1.6*	1.6 \pm 0.7*	5 \pm 5*	8 \pm 5*
<i>Mega TKR patient</i>	2.1 \pm 1.5*	1.5 \pm 0.8*	4 \pm 4*	7 \pm 5*
<i>iNPH patient</i>	3.2 \pm 1.9*	2.3 \pm 1.1*	4 \pm 3*	8 \pm 5*
Total	2.1 \pm 1.5*	1.6 \pm 0.8*	6 \pm 6*	5 \pm 5*

* p < 0.001

Figure 3.3 shows an example of the clusters obtained with the two versions of CIMAP algorithm, considering the Rectus Femoris muscle of a THA patient. The new CIMAP version allows for obtaining 4 and 5 clusters for left and right side respectively, instead of only one cluster per side obtained with the original version. Observing the result of the new CIMAP version is possible to notice that different activation patterns are present: this information was not highlighted by using the original version of the algorithm. Moreover, the clusters resulting by using the new CIMAP present lower values of *Cluster_SD* with respect to the clusters resulting

by using the original CIMAP. Finally, the *CYC_NSC* values are similar for both sides, so no relevant information was lost by using the new version of the algorithm.

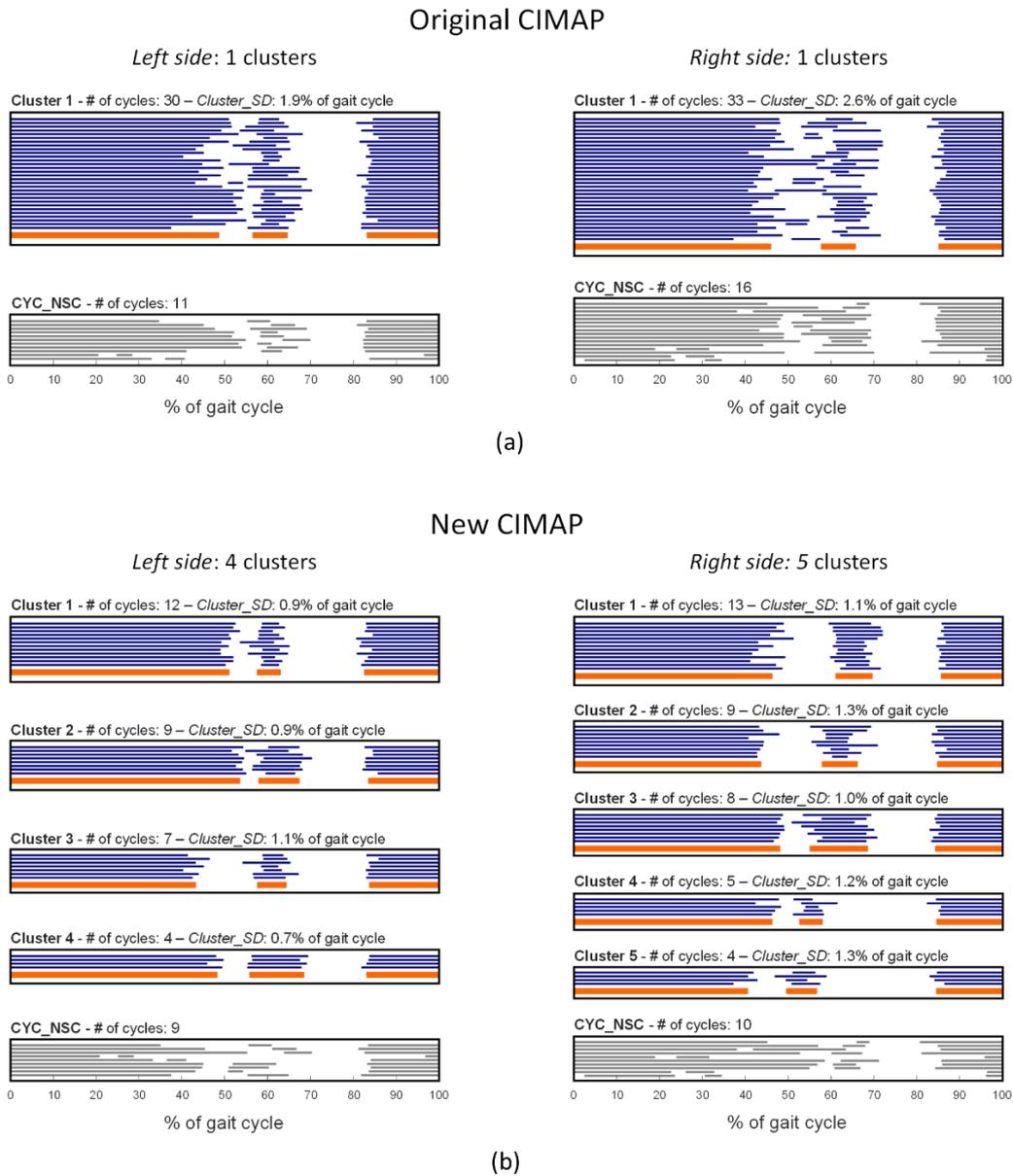


Figure 3.4. Comparison between CIMAP results (original and optimized version).

Example of the clusters obtained with the two versions of CIMAP algorithm for the Rectus Femoris muscle of a THA patient. The *CLUST_SD* values obtained using the new CIMAP version (b) are always lower than those obtained using the original version of the algorithm (a). The *CYC_NSC* values are similar.

3.5 Validation

To validate the new version of the CIMAP algorithm, a new dataset (*Validation_opt*) consisting of both healthy and pathological subject was created. Particularly, gait data of 30 patients aged from 20 to 81 years (15 males/ 15 females, height: 167.5 ± 9.6 cm; weight: 72.0 ± 16.3 kg) were collected:

- 5 healthy young subjects;
- 5 healthy elderly subjects;
- 5 patients after total knee replacement (TKR);
- 5 patients with megaprosthesis of the knee (Mega TKR);
- 5 patients with total hip arthroplasty (THA);
- 5 patients affected by normal pressure hydrocephalus (iNPH).

For each subject, the strides were separated according to their number of activation intervals, obtaining a total of 1070 valid datasets.

The original and the new version of the CIMAP were compared using the average *Cluster_SD* calculated across all the resulting clusters for each dataset. The *Cluster_SD* values and the *CYC_NSC* values for the two versions are shown in Tables 3.4 (mean \pm standard deviation). A paired Student t-test ($\alpha=0.05$) was also performed on each group.

The results obtained on the *Validation_opt* dataset were consistent with the previous founding: the new CIMAP version allows for obtaining a lower intra-cluster variability with respect to the original version, and a significant slightly higher number of cycles is discarded (on average, not more than 4 cycles).

Table 3.4. Comparison of Original and new CIMAP (Validation)

	Cluster_SD (% of g.c.)		CYC_NSC (# cycles)	
	Original CIMAP	New CIMAP	Original CIMAP	New CIMAP
<i>Healthy Elderly</i>	$2.8 \pm 1.8^*$	$1.7 \pm 0.7^*$	$4 \pm 5^*$	$8 \pm 5^*$
<i>Healthy Young</i>	$1.9 \pm 1.6^*$	$1.2 \pm 0.6^*$	$6 \pm 6^*$	$12 \pm 7^*$
<i>THA</i>	$1.9 \pm 1.2^*$	$1.4 \pm 0.7^*$	$3 \pm 4^*$	$7 \pm 6^*$
<i>TKR</i>	$2.5 \pm 1.2^*$	$1.9 \pm 1.0^*$	$4 \pm 4^*$	$9 \pm 7^*$
<i>Mega TKR</i>	$2.8 \pm 1.8^*$	$2.0 \pm 1.3^*$	$3 \pm 3^*$	$7 \pm 5^*$
<i>iNPH</i>	$2.6 \pm 1.6^*$	$2.0 \pm 1.3^*$	$2 \pm 2^*$	$5 \pm 4^*$
<i>Total</i>	$2.4 \pm 1.6^*$	$1.7 \pm 1.0^*$	$4 \pm 4^*$	$8 \pm 6^*$

*p < 0.001

In this chapter the optimized version of the CIMAP algorithm was presented and validated. The results showed that this new version allows for obtaining clusters with lower variability with respect to the original version both analyzing healthy and pathological subjects.

Chapter 4

Principal activation extraction and analysis

4.1 Introduction

In the previous chapters, the CIMAP algorithm was presented and this method is a powerful tool for grouping strides showing similar EMG activation patterns. The results obtained have proved that this method definitively improves the correct interpretation of the acquired signals, providing an organized representation of the most common type of muscle activation patterns of the individual, that may be useful to clinicians to simply identify possible criticism in subject walking.

Moreover, as a result of the clustering procedure, the cluster centroids, both left and right side, are obtained: they summarize the typical activation pattern characteristics of each muscle.

Observing the cluster centroids obtained on healthy subjects during gait, it emerged that, for each specific muscle, some activations are present in almost all the gait cycles and do not belong uniquely to a specific centroid, while other activations are centroid-specific (Figure 4.1a).

It is reasonable to assess that muscle activations that are not centroid-specific are strictly necessary for achieving the specific motor task, and hence they can be defined as *principal activations (PAs)*. The concept of principal activations is complementary to the concept of *secondary activations*, which are activations present only in some strides and have an auxiliary function in motor control (e.g. to provide

a slight correction to muscle activations due to temporary subject distractions or extemporaneous external disturbances) [30].

The concept of principal and secondary activations applied to the analysis of EMG signals may significantly simplify the understanding of muscle contributions to the biomechanics of movement. Principal activations describe the essential contributions of a specific muscle and should be relatively repeatable among similar subjects, while secondary activations are expected to be more variable and possibly related to abnormalities or peculiarities of specific subjects.

In practice, to extract principal activations starting from CIMAP results, at first the prototype of each cluster has to be defined (Figure 4.1a) by coding the cluster centroid as a binary string of 1000 elements: each bit corresponds to 0.1% of gait cycle. A bit equal to 0 means that no muscle activation is observed in the corresponding percentage of gait cycle, while a bit equal to 1 means that muscle activation is observed in the corresponding percentage of gait cycle. Then, the PA of each muscle is obtained as the intersection of the corresponding cluster prototypes (Fig. 4.1b). More specifically, PAs are defined as binary strings of 1000 bits (0 means

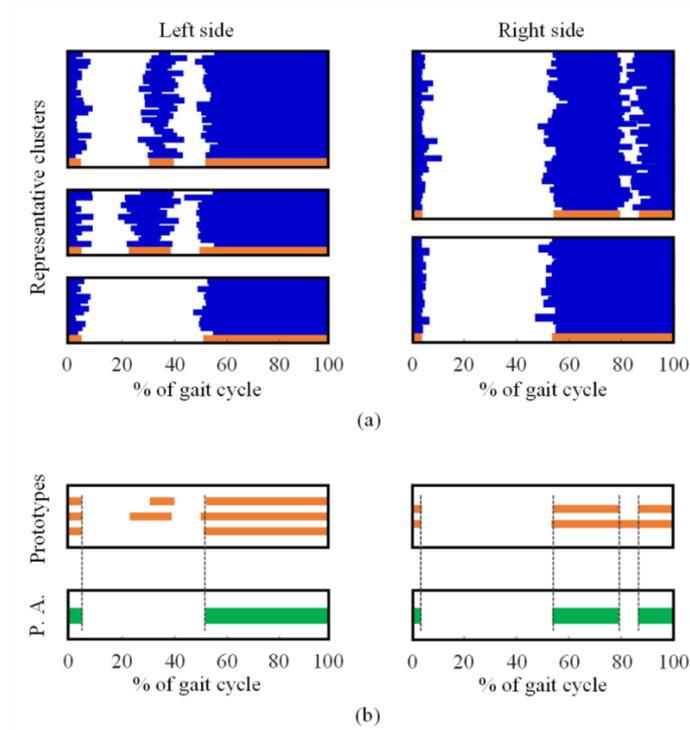


Figure 4.1. Example of *principal activation* extraction.

Example of PA extraction (healthy adult, Tibialis Anterior muscle). In panel (a) the representative clusters are represented: the cluster elements are represented in blue and the cluster prototypes in orange. In panel (b) the principal activations (green) obtained as the intersection of the cluster prototypes are reported.

that at least one prototype has no activation in the specific bit; 1 means that all the prototypes have activation in the specific bit).

Finally, activation intervals lasting less than 3% of the gait cycle are removed, while activation intervals separated by less than 3% of the gait cycle are joined together [38].

4.2 Population characterization using PAs

The extraction of muscle PAs is a new approach to summarize the gait characteristics of a subject, through a single pattern of muscular activation, representative of the activation necessary to correctly perform the gait task.

In this section, the results obtained processing several STEP32 datasets restring the analysis to PAs only, are reported. Particularly, three studies that analyzed group of subjects with homogenous characteristic are performed.

In the first study, a population of healthy school-age children was analyzed through PAs, obtaining normative EMG activation patterns of four lower limb muscles. The results allowed for obtaining a compact information useful to understand the biomechanical contribution of the observed muscles.

In the second study, the PAs analysis was extended to investigate the differences in muscle activation patterns of a healthy population and a group of patients affected by normal pressure hydrocephalus (iNPH). Four lower limb muscles were analyzed and the results showed that focusing the analysis on PAs only, it was possible to clearly identify the main differences in muscle activations between the two groups.

Finally, in the third study, PA analysis was applied to a longitudinal study of patients that underwent total hip arthroplasty surgery (THA). Particularly, muscle activation characteristics were assessed at 3, 6 and 12 months after surgery, and, at each time points, the results on THA patients were compared with the results obtained analyzing a healthy group of control subjects.

In all the studies gait analysis was performed using the STEP32 multichannel recording system. In these studies, all the subjects were equipped bilaterally with (Figure 4.2):

- three foot-switches (size: 10 mm × 10 mm × 0.5 mm; activation force: 3 N) attached beneath the heel, the first and the fifth metatarsal heads of each foot for analyzing the foot-floor contact of both feet;

- one electrogoniometer (with an accuracy higher than 0.5 degrees) attached to the lateral side of the knee joint to record the joint angle during gait;
- several EMG probes attached to the skin over the muscles of interest. The probes used were both fixed and variable geometry probes, depending on the muscle characteristics. The detection modality was single differential and each signal was amplified to best fit the input dynamics of the A/D converter avoiding saturation. Typical gain values ranged from 1000 to 50000. Signals were highpass filtered (10 Hz, two poles), lowpass filtered (450 Hz, six poles) and discretized with a 12 bit A/D converter. The sampling frequency was equal to 2k Samples/sec. The input referred noise was below $1\mu V_{rms}$.

After sensors positioning, subjects walked barefoot for at least 2 minutes, at self-selected speed, back and forth over a straight path ranged from 9 to 15 meters (depending on the study protocol).

All the experimental protocols were conformed to the Helsinki declaration on medical research involving human subjects.



Figure 4.2. Example of sensors positioning.

Example of subject preparation. The subject is equipped bilaterally with three basographic sensors under the foot, one electrogoniometer attached to the lateral side of the knee joint, two fixed geometry probes over Tibialis Anterior and Gastrocnemius Lateralis muscles and two variable geometry probes over Rectus Femoris and Lateral Hamstring muscles.

4.3 Characterization of school-age children population

A dataset consisting of 100 school-age children was retrospectively analyzed, restricting the study to PAs only. In 2010 a contribution [14] describing the role of five limb muscles during walking over the same population was published. For each of the observed muscles - Rectus Femoris (RF), Vastus Medialis (VM), Lateral Hamstrings (LH), Tibialis Anterior (TA) and Gastrocnemius Lateralis (LGS) - five different activation modalities were described.

Restricting the study to PAs only it was expected to obtain a remarkable simplification of the results obtained in the previous study, providing in a novel set of normative EMG activation patterns of simpler interpretation and of higher interest for clinicians [39]. Particularly, at first the global behavior of the population was analyzed and then a more in-depth analysis was performed, analyzing the differences in the EMG activation patterns by dividing children into three groups, according to their age.

4.3.1 Population and gait data acquisition

A population consisting of 100 children (ranged from 6 to 11 years, 51 males and 49 females) was analyzed. Population details are reported in Table 4.1.

Table 4.1. Population details. Details of the 100 children involved in the study (age, height and weight for each group are reported as mean \pm standard deviation).

	Groups (Age range)			
	6.5-8.5 years old	8.5-10 years old	10-11.5 years old	6.5-11.5 years old
<i>Number of children</i>	37	35	28	100
<i>Age (months)</i>	90.4 \pm 7.5	111.5 \pm 5.8	127.6 \pm 4.6	108.2 \pm 16.3
<i>Sex</i>	15M/17F	20M/15F	16M/12F	51M/49F
<i>Height (cm)</i>	125.4 \pm 6.7	136.3 \pm 7.5	140.7 \pm 5.8	133.4 \pm 9.3
<i>Weight (kg)</i>	26.1 \pm 5.1	32.9 \pm 6.8	33.9 \pm 4.9	30.6 \pm 6.7
<i>Number of analyzed strides per child</i>	146 \pm 26	136 \pm 28	115 \pm 34	134 \pm 32

Children with orthopedic or neurological disorders were not included in the study. Subjects walked at self-selected speed for 2.5-minute walk. Parental consent and child assent were obtained prior to the participation in the study.

The muscle activity of Rectus Femoris (RF), Vastus Medialis (VM), Lateral Hamstring (LH), Tibialis Anterior (TA) and Gastrocnemius Lateralis (LGS) of both lower limbs was analyzed (Figure 4.3).



Figure 4.3. Probes positioning.

Representation of the probe positioning on the analyzed muscles: Rectus Femoris (RF), Vastus Medialis (VM), Lateral Hamstring (LH), Tibialis Anterior (TA) and Gastrocnemius Lateralis (LGS).

4.3.2 Data analysis and statistical analysis

After signal segmentation and normalization, gait data were processed using CIMAP algorithm and then muscle PAs were extracted from each subject and coded as strings of 1000 binary bits.

Behavior of the entire pediatric population

To summarize the behavior of the analyzed population the PAs of all the subjects were averaged. More specifically, since PAs were represented as strings of 1000 binary bits, summing the strings point to point across all children, a final string is obtained. Every point of the string has a value from 0 to 100, where 0 means that no child has a PA associated to a certain percent of gait cycle, while 100 means that all children have activation.

Differences among age groups

To analyze possible modifications in walking characteristics due to children development, the population was separated into three age groups: from 6.5 to 8.5 years, from 8.5 to 10.5 years, and from 10.5 to 11.5 years (see Table 4.1). The PAs of each group were represented using the same representation used for representing the PAs of the entire population.

Then left and right data of each group were averaged and the Hotelling's T-square test for two independent samples (significance level: $\alpha = 0.05$) was then applied among the three age groups. More specifically, based on the analysis of the results, the comparison among age groups were performed between 0% and 100% of the gait cycle for VM, LH and TA muscles, while the comparison was limited in the range between 40% and 80% of gait cycle for RF muscle, and between 90% and 100% of gait cycle for LGS muscle.

Ankle control

Several studies in literature have pointed out the importance of the study of the ankle-joint control both in healthy [40] and pathological populations (especially in hemiplegic children [41]-[44]). For this reason, a "more in deep" analysis of the principal activations of ankle plantar- (LGS) and dorsi- (TA) flexors was performed. Particularly the focus of the analysis was on the simultaneous activation of the antagonist muscles (co-contraction [40]) TA and LGS.

4.3.3 Results

Behavior of the entire pediatric population

Figure 4.4 shows the percentage of activation over the entire population of 100 children for the five muscles, observed bilaterally. In order to provide an easier interpretation of those results, the information is reported in two different modalities, for each muscle:

- a grey-coded horizontal bar (frame top): the bar is white when no subject activates the muscle, becomes darker as the number of subjects showing activation at the considered percent of stride increases, and is black when the entire population shows activation of the observed muscle;
- a bar diagram (frame bottom): the diagram shows the same information as the grey-coded horizontal bar, in a different and more quantitative

way. More specifically, each bar corresponds to a specific percent of the gait cycle and its height is proportional to the percent of subjects showing muscle activation.

The horizontal axis reports the percentage of the gait cycle and refers to all the graphs.

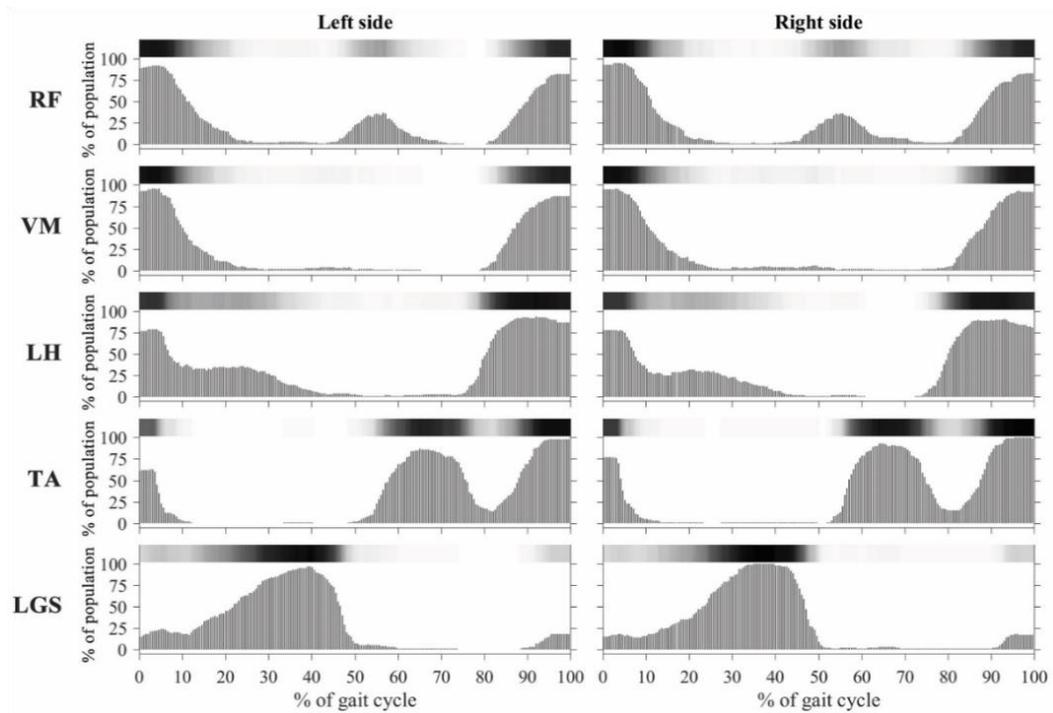


Figure 4.4. Activation representation of 100 school-aged children.

Gray-coded bar (frame top) and bar diagram (frame bottom) representing, for each percent of the stride, the percentage of children in which a PA is observed. Results are reported for RF, VM, LH, TA and LGS muscles, left and right leg separately.

Differences among age groups

Figure 4.5 shows the percentage of activation over three sub-populations (different for subject age) for the five analyzed muscles, bilaterally. In this case, to facilitate comparison among sub-populations, results are reported using the bar diagrams only. Each bar corresponds to a specific percent of the gait cycle stride and its height is proportional to the percentage of children, in the specific sub-population, showing muscle activation.

Comparing the principal activations among age groups and the Hotelling's T-square test result (Table 4.2), several observations can be made:

- a) For RF muscle, between 40% and 80% of gait cycle, principal activations are significantly different among all age groups.
- b) For VM muscle, between 0% and 100% of gait cycle, principal activations are significantly different only between Group 1 and 3.
- c) For LH muscle, between 0% and 100% of gait cycle, principal activations are significantly different between Group 1 and 3, and between Group 2 and 3.
- d) For TA muscle, between 0% and 100% of gait cycle, principal activations are not significantly different among age groups.
- e) For LGS muscle, between 90% and 100% of gait cycle, principal activations are significantly different between Group 1 and 2.

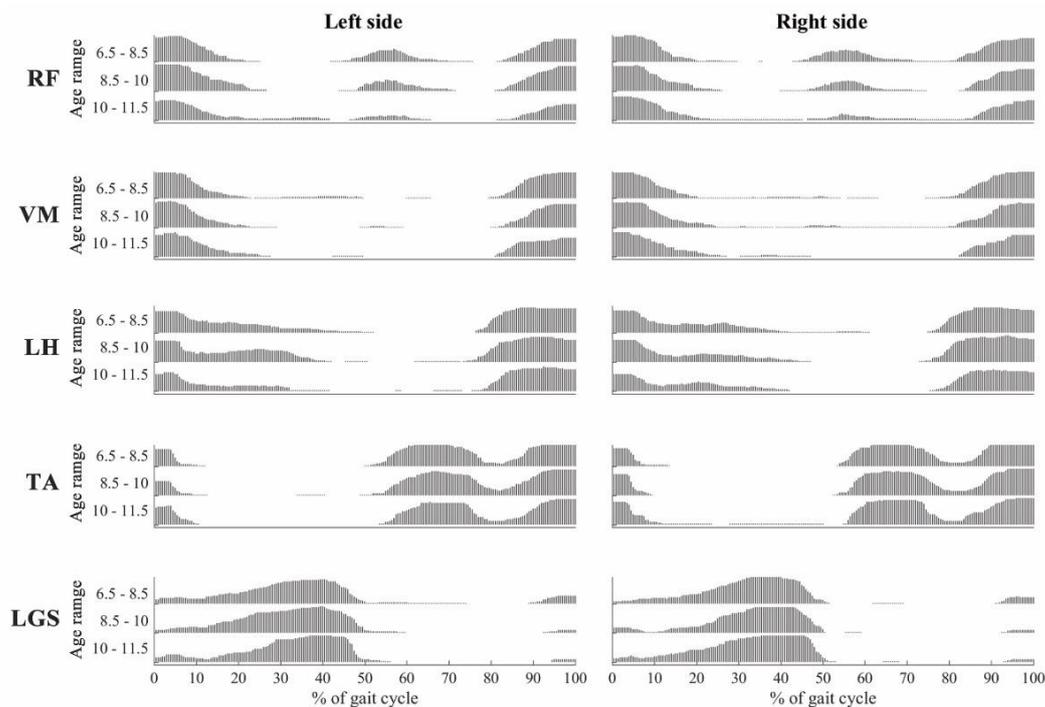


Figure 4.5. Activation representation of three sub-population of the 100 school-aged children.

Bar diagrams representing, for each percent of gait cycle, the percentage of children in which a principal activation is observed in each sub-population. Results are reported for RF, VM, LH, TA and LGS muscles, left and right leg separately.

Table 4.2. Comparison among age groups: results of the Hotelling’s T-square test (p-values).

	Muscles				
	RF	VM	LH	TA	LGS
<i>Group 1 vs. Group 2</i>	0.0023	0.17	0.99	0.68	<0.0001
<i>Group 1 vs. Group 3</i>	<0.0001	0.0074	<0.0001	0.94	<0.0001
<i>Group 2 vs. Group 3</i>	<0.0001	0.19	<0.0001	0.73	0.38

*Group 1: from 6.5 to 8.5 years, Group 2: from 8.5 to 10 years, Group 3: from 10 to 11.5 years. p-values < 0.05 are highlighted in bold.

Ankle control

By analyzing the PA of ankle plantar- (LGS) and dorsi- (TA) flexors it emerges that approximately 25% of the population activates LGS following the initial contact (0-10% of gait cycle) where TA muscle is active in approximately 60% (left leg) and 75% (right leg) of the population. 18 subjects show co-contraction of TA and LGS on the left limb and 15 subjects on the right limb. Before the initial contact (90-100% of GC), approximately 18% of the population shows an anticipate activation of plantar flexors and consequently a co-contraction of LGS and TA. Moreover, Figure 4.5 shows that the anticipated activation of LGS is progressively less evident with increasing age.

4.3.4 Discussion

Among all, comparing the results provided in the study published in 2010 [14] (Figure 4.6) with the results obtained analyzing PAs (Figure 4.4) it is evident that a this new approach should simplify the analysis. In [14] the presence of multiple activation modalities for each muscle, determined a complex picture for the overall

motor patterns of children during gait and their interpretation. On the contrary, considering principal activations only, allows for simplifying the analysis and providing data simpler to be interpreted that may be more useful in clinical practice.

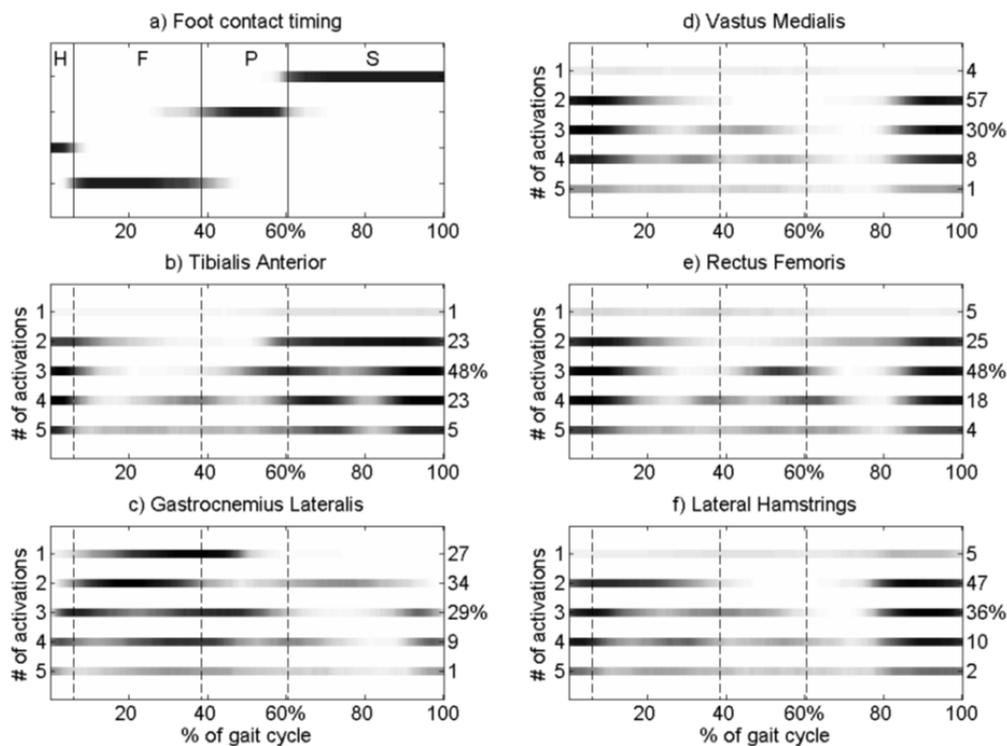


Figure 4.6. Results found in [14]

“Horizontal bars are grey-level coded; black: condition observed for all children, white: condition never met. a) Foot contact timing over the population (4-level signal): H = heel contact; F = Flat foot contact; P = Push off; S = Swing. Muscle activation onset and offset instants over the population for b) TA, c) GL, d) VM, e) RF and f) LH, as % of gait cycle, for five different modalities with 1, 2, 3, 4 and 5 activations, respectively. On the right-hand side of each plot: percentage frequency of each modality. The H, F, P, S phases are shown superimposed, delimited by dashed vertical lines.”

Behavior of the entire pediatric population

In terms of hip and knee control, in [14] results showed that the RF muscle in the 90% of analyzed strides showed 2, 3 or 4 activation modalities (Figure 4.5e). Restricting the analysis to PAs, it emerged that approximately 35% of the children present a 3-activation modality and approximately 60% relies on two activations (Figure 4.3). Particularly, in case of 2 activation modality, the muscle is activated around the initial contact, to avoid excessive knee flexion during the weight acceptance phase. The third activation, when present, extends from 45% to 70% of the gait cycle and it is aimed at stabilizing the trunk before toe-off.

Considering the VM muscle, in [14] results showed that the 57% of the gait cycle presented 2 activations intervals, while 30% and 8% of gait cycle presented 3 and 4 activation intervals, respectively (Figure 4.5d). Analyzing only PAs it emerged that over the 90% of children used a 2-activation modality (Figure 5.3), activating the muscle around the initial contact, mainly for stabilizing patella in preparation and during the weight acceptance phase.

Similarly, in [14] the LH muscle showed 2-, 3- or 4-activation modalities respectively in 47%, 36% and 10% of strides, while restricting the study to PAs a large percentage of the population uses a 2-activation modality (Figure 4.3). More specifically over 90% of the children show activation from 85% to 95% of the gait cycle. Since in this interval the knee is at the end of its extension phase, this is an eccentric contraction aimed at slowing down the joint extension in preparation to the initial contact. Notice that approximately 75% of the population shows a simultaneous activation of RF and LH around the initial contact, aimed to the stabilization of knee and hip joints in preparation to the weight acceptance phase. Only 25% - 30% of the population shows LH activity from 10% to 30% of the gait cycle, where RF is not active, probably to extend the hip during mid stance.

Considering ankle control, the results reported in [14] were consistent with those reported in this study. In [14] the TA muscle showed 2-, 3- or 4-activation modalities respectively in 23%, 48%, and 18% of strides. Considering PAs only, the TA muscle shows a 3-activation modality in approximately 85% of the population and a 2-activation modality in slightly less than 15% of the population. The activation of the TA begins approximately at 55% - 60% of the GC, just before toe-off. During the swing phase approximately 85% of the population deactivates the muscle around 82% of the GC. Almost the entire population shows activity just before initial contact and 75% of subjects keep the muscle active up to 6% of the gait cycle to control the first heel rocker. After 15% of GC no subject shows TA activity until toe-off.

Differences among age groups

Observing Figure 4.5, that reports the results related to the observed muscles divided into three age groups, two significant observations emerged: first, with increasing age, it is observed that the activity of the RF tends to disappear during push-off; secondly, also the activation of the LGS muscle preceding the initial contact (90% - 100% of GC) tends to be less frequent in older children.

During push-off the RF muscle is activated during knee flexion, thus acting like a hip flexor. The goal of hip flexion during push-off is to control the position of the trunk to move the center of gravity forward. However, with increasing age, it is

observed that the LGS muscle is activated more frequently between 35% and 45% of GC, thus improving its propulsive effect. Consequently, it may be concluded that with increasing age the advancement of the body tends to rely more on the propulsive effect of the plantar flexors and therefore the need to move forward the center of gravity decreases, thus making the body more stable during the progress. This may be interpreted as a sign of gait maturation.

Observing the activity of the TA muscle, it emerges that from 90% to 100% of the GC the muscle is always active: this activity is aimed at obtaining plantar extension and preparing the forefoot for contact between the ground and the foot. The simultaneous activation of the LGS muscle in this phase of GC aims at stiffening the ankle joint to obtain a better control of the heel contact phase. Figure 4.5 shows that the LGS activation just before heel contact is present in approximately 30% of children belonging to the younger group and becomes less frequent in older children. A possible interpretation of this finding is that the need for stiffening the ankle joint just before heel contact decreases with increasing age, thus demonstrating more natural ankle control in older children

Ankle control

Even if it is generally accepted [1] that around the heel contact plantar flexors and extensors should not be co-activated, the results obtained in this study shows that approximately 16%-18% of children presents a co-activation of LGS and TA muscles around the initial contact, confirming the founding of a previous study [45]. Around the heel contact, among the observed muscles, only the TA muscle is expected to be active, but this is not true for approximately 22% of the analyzed population. More specifically, before heel contact, approximately 18% of the population activates the LGS muscle in co-activation with the TA muscle; after heel contact the percentage of children that activates the LGS muscle is slightly higher (approximately 25%), but approximately 8% of the population does not activate TA and LGS contemporarily. It may be concluded that, in general, the co-contraction of plantar flexors and extensors around the initial contact is not so uncommon as usually assumed. With the increasing of the age, the percentage of children presenting co-activation of TA and LGS before the initial contact decreases considerably, while the frequency of co-contraction of plantar flexors and extensors following the initial contact (0% - 10% of GC) does not depend on age.

4.3.5 Conclusions

In conclusion, analyzing the muscle activation intervals during gait considering PAs only, makes it easier to understand the biomechanical contribution of the observed muscles and it allows describing muscle activity in an easier form that may be more useful for clinical practice. Particularly, in this study, it was possible to identify and describe phenomena related to gait maturation that were less evident when all the muscle activations were considered.

4.4 Assessment of muscle activation differences between idiopathic normal pressure hydrocephalus patients and healthy controls

In the previous section, PA analysis was applied for describing the global behaviour of a population of 100-school age children. The results obtained restricting the study to PAs only provided both qualitative and quantitative information about the population. However, in a typical clinical scenario, it is often required not to analyse a single dataset, but to compare datasets from:

- a) a single patient or a patient cohort with respect to a control group [46];
- b) single patient or a patient cohort before and after a medical treatment [47];
- c) repeated observations of a single patient or a patient cohort over long periods of time (follow-up or longitudinal study) [48].

However, the problem of dataset comparison is very challenging due to the EMG stride-to-stride variability already documented. To cope with this problem, the application of the CIMAP algorithm and the PA extraction may be a powerful tool for the aggregation of EMG results of huge dataset, and it may allow an easy comparison between different datasets.

This section deal with the problem of comparing datasets from a patient cohort with respect to a control group. Particularly, the patient cohort is composed of subject affected by idiopathic normal pressure hydrocephalus (iNPH).

Normal pressure hydrocephalus (NPH) is an abnormal buildup of cerebrospinal fluid (CSF) in the brain's ventricles, or cavities. This neurological disorder is characterized by gait disturbance, dementia and urinary incontinence [49]. Particularly, gait alteration is usually the first sign to appear and is considered the most important symptom [50], [51], [52].

EMG gait data of 30 patients affected by normal pressure hydrocephalus and 30 healthy controls were extracted from STEP32 database and analyzed.

Aim to this study is to analyze PAs of both patients and controls to investigate the differences in muscle activation patterns of the two population.

4.4.1 Population and gait data acquisition

Gait data acquired from 30 iNPH patients (ranged from 50 to 86 years, 25 males and 5 females) and 30 matched controls (ranged from 40 to 75 years, 15 males and 15 females) were extracted from STEP32 database and retrospectively analyzed. Population details are reported in Table 4.3.

Table 4.3. Population details. Anthropometric characteristics of the Sample Populations (iNPH and Controls). Age, height and weight for each group are reported as mean \pm standard deviation.

Anthropometric Characteristics	iNPH Patients (N = 30)	Controls (N = 30)
<i>Age (years)</i>	73.8 \pm 8.6	61.9 \pm 9.5
<i>Height (cm)</i>	167.6 \pm 7.6	170.7 \pm 9.1
<i>Weight (kg)</i>	72.0 \pm 11.0	69.4 \pm 11.4

All patients showed a short-stepped “magnetic” gait, cognitive disturbances and, in many cases, urinary incontinence [5].

The muscle activity of Lateral Hamstrings (LH), Rectus Femoris (RF), Gastrocnemius Lateralis (LGS) and Tibialis Anterior (TA) of both lower limbs was analyzed (Figure 4.7).

4.4.2 Data analysis and statistical analysis

The H, F, P and S gait phases of each subject enrolled in the study were extracted using the routines for gait analysis included in the STEP32 software. The strides were segmented in separate gait cycles and only HFPS were extracted and normalized to time duration [21]. The onset-offset timings of each EMG activation were detected using a double-threshold statistical detector [22].

The CIMAP algorithm was applied to the entire dataset and muscle PAs were obtained and coded as binary strings of 1000 bits.



Figure 4.7. Probes positioning.

Representation of the probe positioning on the analyzed muscles: Lateral Hamstrings (LH), Rectus Femoris (RF), Gastrocnemius Lateralis (LGS) and Tibialis Anterior (TA).

The Lilleforst test was applied to assess the normality of the distribution consisting of the values of H, F, P and S phases, for both the groups. The differences in gait phase duration between patients and controls were then assessed using Student's *t*-tests ($\alpha=0.05$) when the distribution resulted normal, and using Wilcoxon non parametric tests ($\alpha=0.05$) when at least one of the two tested distribution resulted not-normal.

Since, as already documented in literature [5], significant differences were observed between the phase durations of iNPH patients and controls, H, F, P and S phases of iNPH patients were time-normalized with respect to the phase durations of controls.

4.4.3 Results

For each lower limb, an average of 68 ± 19 HFPS gait cycles were collected for iNPH patients and 77 ± 12 for controls. The values of gait phases time duration for left and right side are reported in Table 4.4, separately on the two groups.

In Table 4.5 are reported the results of the Lilleforst test, for the H, F, P and S phases separately, both groups. As it emerges from the table, the distribution of values related to the time duration of H phase (left side) of both groups, resulted not-normal, as well as on the right side for iNPH patients. Moreover, on the control group, the distribution of values related to the time duration of P phase (right side) resulted not-normal.

Table 4.4. Gait phase time duration. Values are reported as mean \pm standard deviation.

	Left side		Right side	
	Controls	iNPH patients	Controls	iNPH patients
<i>H</i>	5.6 \pm 2.2	6.4 \pm 3.9	5.1 \pm 1.5	6.2 \pm 4.5
<i>F</i>	32.8 \pm 4.1	43.8 \pm 7.0	32.8 \pm 5.0	42.9 \pm 7.0
<i>P</i>	20.3 \pm 4.6	15.5 \pm 5.1	20.4 \pm 5.3	16.0 \pm 4.7
<i>S</i>	4.4 \pm 3.1	34.4 \pm 3.5	41.6 \pm 2.1	34.9 \pm 4.8

Table 4.5. Lilliefors test results. Not-normal distribution are identified by p-values $<$ 0.05.

	p-values			
	Left side		Right side	
	Controls	iNPH patients	Controls	iNPH patients
<i>H</i>	0.04	0.007	0.05	0.004
<i>F</i>	0.5	0.3	0.3	0.3
<i>P</i>	0.08	0.3	0.04	0.3
<i>S</i>	0.22	0.2	0.09	0.5

*p-values $<$ 0.05 are highlighted in bold.

The results of Student t-tests and Wilcoxon tests reported in Table 4.6 highlight significant differences in phase durations between controls and iNPH patients for F, P, and S phases, both sides.

Table 4.6. Student t-test and Wilcoxon test results.

	p-values	
	Left side	Right side
<i>H</i>	0.4	0.5
<i>F</i>	<0.001	<0.001
<i>P</i>	<0.001	0.001
<i>S</i>	<0.001	<0.001

*p-values $<$ 0.05 are highlighted in bold.

In Figure 4.8 the percentage of subjects having PA at each percent of gait cycle is shown separately for the four analysed muscles (TA, LGS, RF and LH).

For each muscle, the results obtained with (panel a, c, e and g) and without (panel b, d, f and h) gait phase normalization are reported.

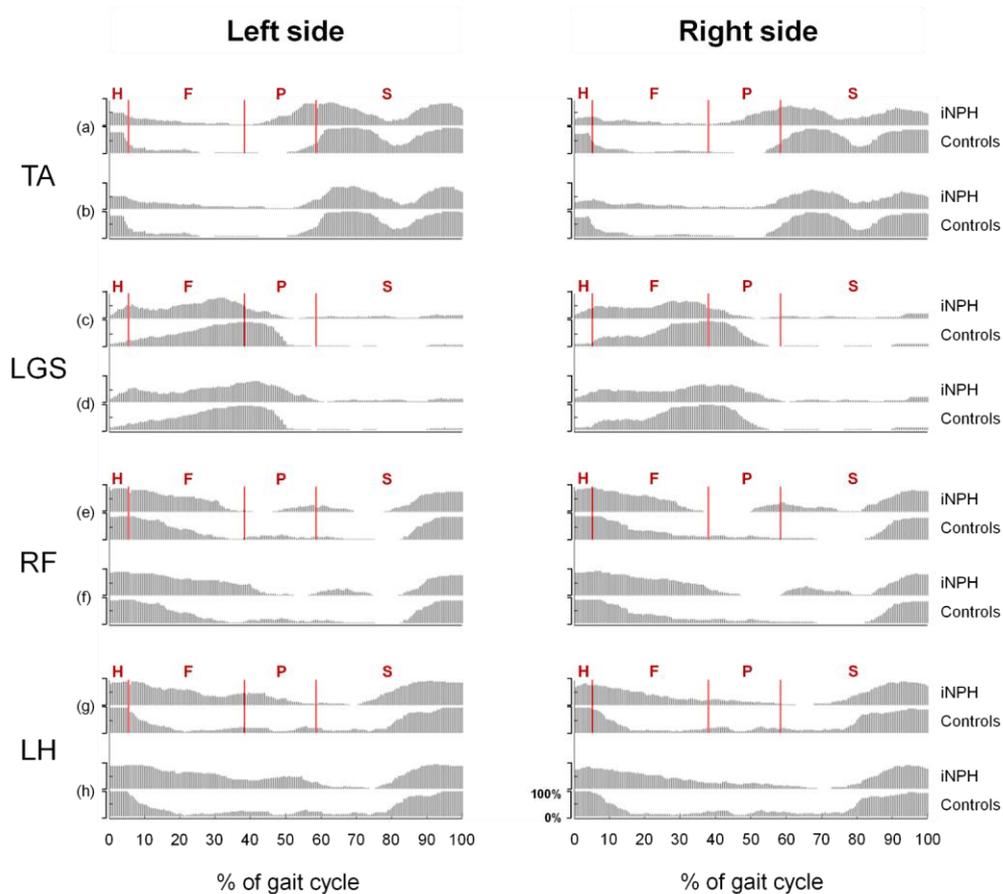


Figure 4.8. Activation representation of iNPH patients and controls.

Barplots representing, at each percent of gait cycle, the percentage of subjects showing a principal activation, for both iNPH patients and controls. For each muscle, the results obtained with (panel a, c, e and g) and without (panel b, d, f and h) gait phase normalization are reported. Muscles: Tibialis Anterior (TA), Gastrocnemius Lateralis (LGS), Rectus Femoris (RF) and Lateral Hamstring (LH).

4.4.4 Discussion

Observing Figure 4.8 it can be noticed that, overall, qualitatively the same behaviour is exposed on left and right side, for both the groups. Hence, in the following, the results obtained will be discussed separately for each muscle, but all the observation will be referring to both sides.

Tibialis Anterior

From both panel *a* and *b* it emerges that, in the first 6-7% of the gait cycle, there is a difference in the number of subjects that activate the muscle, suggesting that around 50% of iNPH patients does not activate TA muscle at the beginning of the cycle. Observing panel *b*, no other significant differences emerged. On the contrary, the results obtained after gait phase normalization (panel *a*), highlights a marked anticipation of the onset timing during P phase.

Gastrocnemius Lateralis

The behaviour of iNPH patients and controls results qualitatively different during H and F phases (panel *c* and *d*): only a small percentage of controls activate the LGS muscle at the beginning of the gait cycle, while the great majority of subjects present an onset timing starting from the middle of P phase. Instead iNPH patients expose a more variable behaviour in the onset timing of the muscle. Moreover, observing panel *d* of Figure 4.8 it emerges a slightly delay in the offset timing of the muscle activation during F phase. On the contrary, the result reported in panel *c* highlights an anticipation of offset timing for the majority of iNPH patient with respect to controls.

Rectus Femoris

Small differences can be observed between patients and controls for RF muscle. Particularly, a delay in offset timing during the final part of F phase and the beginning of P phase can be observed from panel *f*. This difference is not highlighted when activations are normalized to control gait phase duration (panel *e*), but in this case a slightly anticipation of the onset timings at the end of gait cycle is observed for iNPH patients with respect to controls.

Lateral Hamstring

Qualitatively, the behavior of iNPH patients results quite similar both with and without gait phase normalization (panel *g* and *h*). Particularly it emerges a prolonged muscle activity during F and P phases with respect to control group.

For all the muscle, except for LH, the gait phase normalization allowed for obtaining a different kind of information with respect to the standard approach. Analyzing both results it was possible to extract two kind of information. At first, when activations are not normalized the results referred to the whole gait cycle and they underlined global difference between patients and controls. At second, applying the gait phase normalization, it was possible to underline difference between the two

groups referring to the specific gait phase, bringing out different kind of muscle activation strategies to perform the walking task.

4.4.5 Conclusion

In this section, the PA analysis was applied to both healthy subjects and iNPH patients and it has proved to be a powerful tool to compare muscle activation patterns of the two groups. Moreover, the concept of gait phase time normalization was introduced, and it has revealed to be a useful approach to extract information, especially when significant differences are present in gait phase duration of the groups of interest.

The presented approach can be easily extended for the analysis of other pathologies, becoming a promising tool for clinicians to identify the main differences between healthy and pathological subject and to implement effective rehabilitation protocol.

4.5 Longitudinal assessment of muscle function in patients with Total Hip Arthroplasty

In this section, gait data acquired from 20 patients with Total Hip Arthroplasty (THA) at 3, 6 and 12 months after surgery and 20 matched control subjects are analysed, in order assess the efficacy of PA analysis in a longitudinal study. In 2014 a contribution [4] describing the role of five limb muscles during walking over the same population was published. Restricting the study to PAs only, it was expected to compare in a more compact way the datasets, providing a robust, synthetic and clear representation of EMG gait data that might be highly appreciated by clinicians [53].

Total Hip Arthroplasty is the orthopedic surgical replacement of the hip joint with an artificial prosthesis and it is usually proposed to patients that previously underwent other treatment options that have failed or did not provide adequate pain relief. The procedure consists in the removal of damaged bone and cartilage, through an anterior or posterior approach, and in their replacement with prosthetic components [54]. After the surgery procedure, alterations in locomotor activity may persist and often a follow-up that involves specialized physiotherapists may be required [55].

4.5.1 Population and gait data acquisition

A population consisting of 20 THA patients (ranged from 49 to 79 years, 9 males and 11 females) and 20 matched controls (ranged from 57 to 74 years, 11 males and 9 females) was analyzed. Population details are reported in Table 4.7.

Table 4.7. Population details. Anthropometric characteristics of the sample populations (THA and Controls). Age, height and weight for each group are reported as mean \pm standard deviation.

Anthropometric Characteristics	Hip Prosthesis Patients (N = 20)	Controls (N = 20)
<i>Age (years)</i>	66.1 \pm 7.2	65.4 \pm 5.1
<i>Height (cm)</i>	168.7 \pm 10.5	169.8 \pm 9.4
<i>Weight (kg)</i>	77.0 \pm 13.3	69.0 \pm 12.2

All patients originally suffered from hip osteoarthritis and, after surgery, they received the same 2 months rehabilitation protocol [4]. The patients' outcome was assessed at 3, 6, and 12 months after surgery. The muscle activity of Gluteus Medius (GMD), Lateral Hamstring (LH), Rectus Femoris (RF), Gastrocnemius Lateralis (LGS) and Tibialis Anterior (TA) of both lower limbs was analyzed (Figure 4.9).

4.5.2 Data analysis

Using the dedicated routines included in the STEP32 software, the H, F, P and S gait phases were obtained, for each lower limb. The signal was then segmented in separate gait cycles and only cycles showing the normal sequence of gait phases (HFPS) were extracted and normalized to time duration [13]. The onset-offset timing of each EMG activation was detected using a double-threshold statistical detector [56].

The CIMAP algorithm was applied to the entire dataset; then, for THA patients, each resulting cluster was separated into strides belonging to prosthetic and sound side; for healthy controls, clusters were separated into left and right side. For all subjects, muscle PAs were obtained and coded as binary strings of 1000 bits.

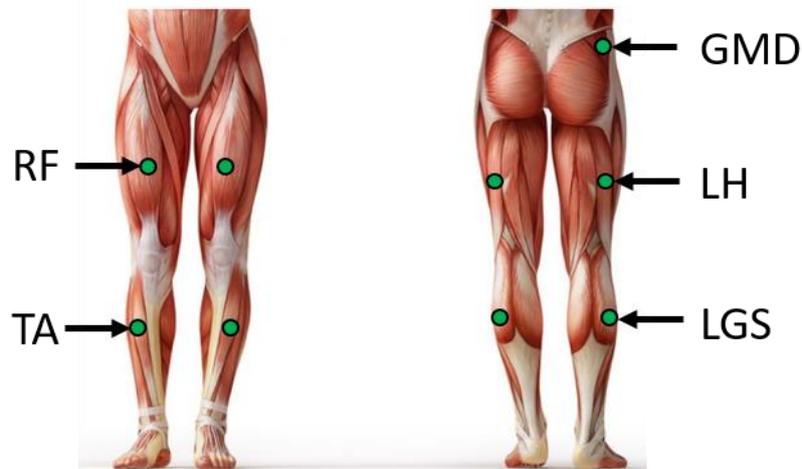


Figure 4.9. Probes positioning.

Representation of the probe positioning on the analyzed muscles: Gluteus Medius (GMD), Lateral Hamstring (LH), Rectus Femoris (RF), Gastrocnemius Lateralis (LGS) and Tibialis Anterior (TA).

Muscle PAs of each analyzed groups (THA patients at 3 time points and controls) were averaged across all subjects belonging to the same group.

For THA group, a string with values that lies between 0 and 100% was obtained, at each time point, both on prosthetic and sound side. For control group, PAs of both left and right side were averaged to obtain a unique reference to compare both prosthetic and sound side of THA patients.

As mentioned before, the gait cycles were time normalized with respect to the gait cycle duration and expressed as a percentage of the gait cycle (% G.C.). However, since significant differences were observed between gait phase durations of THA patients and controls [4], H, F, P and S phases of THA patients were time-normalized with respect to that of controls (averaging left and right side).

4.5.3 Results

Analyzing the entire dataset, an average of 72 ± 15 and 76 ± 13 HFPS gait cycles were collected for THA and control group, respectively. In Figure 4.10 the percentage of subjects showing PA at each percent of gait cycle is shown. THA gait phases were normalized with respect to the mean phase duration of controls (H: 5.0

± 1.1 %GC, F: 34.3 ± 3.4 %GC, P: 19.4 ± 2.8 %GC, S: 41.3 ± 2.1 %GC). The information is reported as barplot, separately for the five analysed muscles.

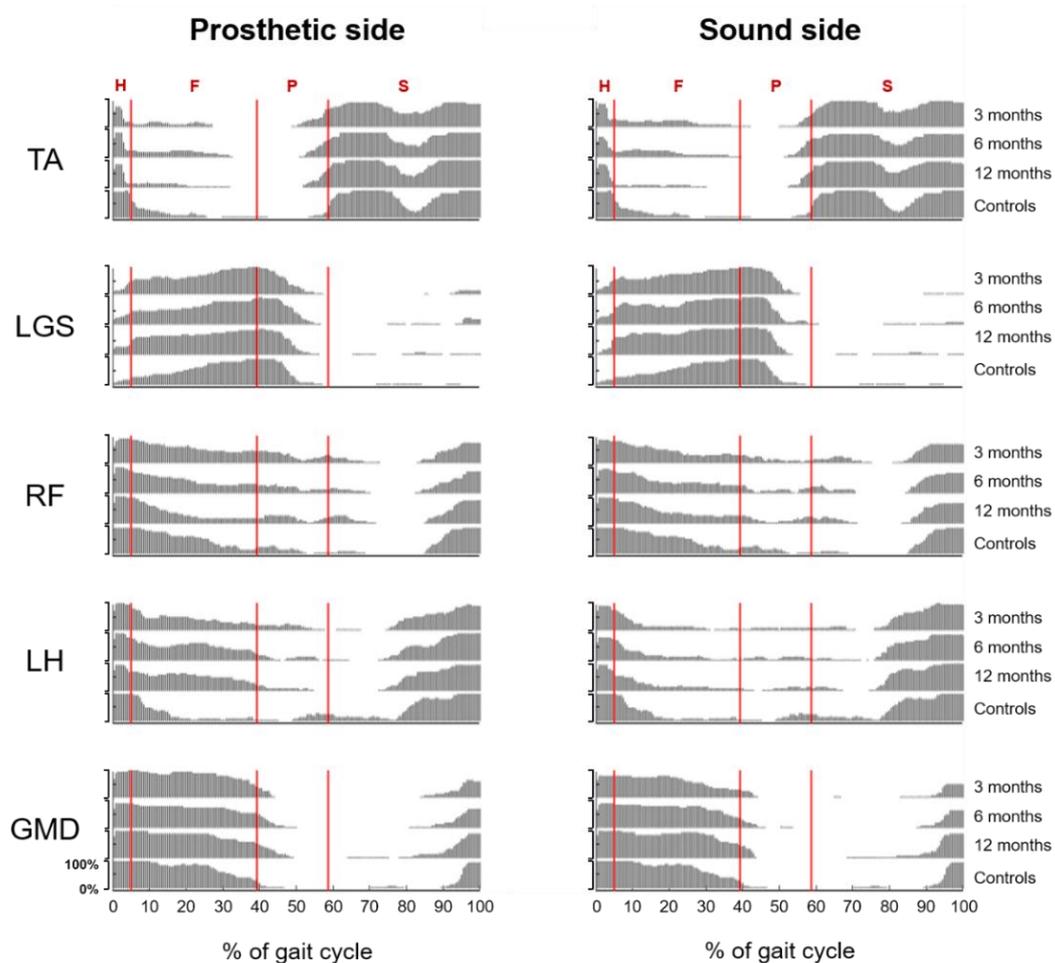


Figure 4.10. Activation representation of THA patients at 3, 6 and 12 months after surgery, and controls.

“Barplots representing, at each percent of gait cycle, the percentage of subjects showing a principal activation. For each muscle, the barplots of THA group at 3, 6 and 12 months after surgery (both prosthetic and sound side) and control group are represented. TA: Tibialis Anterior, LGS: Gastrocnemius Lateralis, RF: Rectus Femoris, LH: Lateral Hamstring and GMD: Gluteus Medius.” [53]

4.5.4 Discussion

Considering the TA muscle, it can be noticed that the large majority of THA patients show an anticipation of offset timing during H phase with respect to control group. This behaviour is present both on prosthetic and sound side and it doesn't change significantly during the follow-up. Furthermore, observing the final part of

the P phase, it can be observed that THA patients already show muscle activity while in controls group this activity appears later, during the S phase. This behaviour was observed clearly on prosthetic side and, to a minor extent, also in the sound side.

Observing the LGS muscle, the only remarkable difference between THA patients and controls is that a larger percentage of THA patients activate the muscle in the first part of F phase, on both sides.

The LH muscle is the only muscle that presents relevant differences in the behaviour of prosthetic and sound side in THA patients. More specifically, for the prosthetic side, the main difference is evident during P phase, where a greater percentage of THA patients show PA with respect to controls. For the sound side, the main difference is evident during the H phase, where a smaller percentage of THA patients show PA with respect to controls.

Finally, for what concern the GMD and RF muscles, no relevant differences can be appreciated between THA and control groups.

4.5.5 Conclusion

In this section, the problem of comparing EMG gait data among populations (and among different time points) was addressed using PAs extracted through the CIMAP algorithm. The information that was obtained is complementary to that of a previous works in many aspects [4] and provides a compact description of the muscle activation patters in the analysed populations.

In conclusion, the proposed methodology based on muscle Pas, allowed for comparing, in a compact way, the results from a longitudinal study conducted on patients after hip replacement surgery. This methodology can be easily extended to the analysis of other longitudinal studies or can be applied to compare datasets related to different population showing homogenous characteristics.

Chapter 5

EMG-based indices for gait data interpretation

5.1 Introduction

In the first chapters, the CIMAP method was presented. This method has shown to be able to group strides with similar EMG-activation patterns, and to highlight the subject's most representative activation patterns exposed during a walking session.

Then, the concepts of principal and secondary activations have been introduced to distinguish those activations necessary to correctly perform the gait task (principal), to those activations present only in some strides (secondary), aimed at the biomechanical stabilization of the body under particular environmental conditions. Some examples have been presented to show how the analysis of the principal activations is useful for the synthetic characterization of a population and how this analysis can be extended to the study and comparison of different populations (i.e. iNPH vs Controls) or to analyze the follow-up of patients that underwent a surgical procedure (follow-up of THA at 3, 6 and 12 months after surgery).

In this chapter, some tools aimed to the interpretation extracted from principal activations will be presented. Particularly, two indexes based on muscle principal

activations will be introduced: the first one aims to quantify the asymmetry in muscle activations during walking and the second one aims to quantify the differences in muscle activation patterns of a specific subjects with respect to a population of reference.

5.2 EMG Asymmetry Index

Gait asymmetry is generally defined as the different behavior of the lower limbs during locomotion. A quantitative and objective measurement of the asymmetry level in lower limb during walking can provide significant insight about the control of walking [57].

Particularly, in clinical scenario, the automatic identification and quantification of gait asymmetry is very important. At first, gait asymmetry may be associated with a number of negative consequences, such as difficulty in balance control, inefficiency, risk of musculoskeletal injury to the non-paretic lower limb and loss of bone mass density in the paretic lower limb [58], [59]. Moreover, pronounced asymmetry levels have been associated with pathological conditions such as cerebral palsy, stroke, osteoarthritis, and knee and hip arthroplasties. For these reasons the quantification of gait asymmetry may have an important role in guiding the clinician's treatment decisions.

Despite the potential importance that the definition of unique standards to quantify the asymmetry may have, currently there are no commonly accepted standard for either the method used to calculate gait symmetry or the gait parameter to assess [57]. In [60] authors present a kind review of the existing quantitative symmetry assessment methods for gait that has been summarized in Figure 5.1.

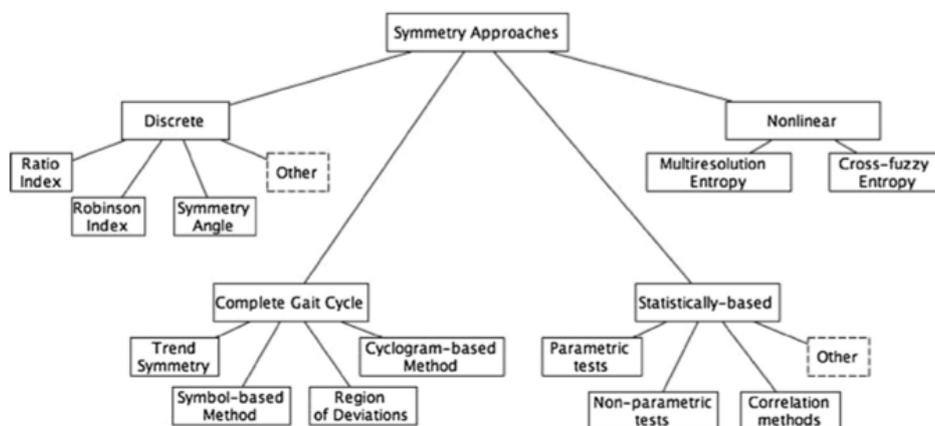


Figure 5.1. Overview of classification of symmetry assessment approaches [60].

As it emerges from [60] and from other studies found in literature, several different gait asymmetry indices can be found [61]-[64]. Most of the proposed indices based the asymmetry quantification on spatio-temporal [58], [65] or joint kinematics parameters [66], [67] and only a few studies addressed the problem of defining an asymmetry index based on sEMG signals [68], [69].

The purpose of this section is then to propose a robust sEMG asymmetry index, based on muscle PAs, for assessing muscle-activation asymmetry in cyclic movements [70]. At first the definition of the proposed index is provided; then the index is used to assess the asymmetry level of four muscles in healthy controls, and in neurological and orthopedic patients. Finally, a critical comparison is also provided with other EMG asymmetry indices found in literature [68], [69].

5.2.1 Population and gait data acquisition

A population consisting of both healthy and pathological subject was analyzed. Gait data of 114 subjects were extracted from STEP32 database:

- 30 healthy subjects (controls): 10 elderly, 10 adults and 10 children.
- 49 orthopedic patients, including 19 adults with megaprosthesis of the knee after bone tumor resection (*Mega TKR*), 10 elderly subjects with Total Knee Replacement (*TKR*) and 20 elderly subjects with Total Hip Arthroplasty (*THA*) evaluated at 3, 6 and 12 months after surgery.
- 35 neurological patients, including 25 children with Winters' type I and II hemiplegia (*Hemiplegic Children*) after cerebral palsy, and 10 elderly subjects with idiopathic Normal Pressure Hydrocephalus (*iNPH*).

Population details are reported in Table 5.1.

The muscle activity of Tibialis Anterior (TA), Gastrocnemius Lateralis (LGS), Rectus Femoris (RF) and Lateral Hamstrings (LH) of both lower limbs was recorded and analyzed (Figure 5.2).

Subjects walked barefoot, at self-select speed, back and forth over a straight path (walkway length: from 7 to 15 m, depending on the study protocol from which they were extracted), for at least 150s.



Figure 5.2. Probes positioning.

Representation of the probe positioning on the analyzed muscles: Tibialis Anterior (TA), Gastrocnemius Lateralis (LGS), Rectus Femoris (RF) and Lateral Hamstrings (LH).

5.2.2 Definition of the sEMG asymmetry index

For each subject, the CIMAP algorithm was applied and then left and right muscle PAs were extracted. Finally, an index based on PAs (*EMG_ASYM_INDEX*) was defined to quantify the muscle-activation asymmetry. For each muscle, the *EMG_ASYM_INDEX* was calculated according to (5.1):

$$EMG_ASYM_INDEX = \sum_{i=1}^N \left| \frac{R_i - L_i}{N} \right| \times 100\% \quad (5.1)$$

where *R* and *L* identify the binary strings corresponding to the PAs of right and left sides respectively and *N* is the number of elements used for representing the principal activations (*N*=1000).

Notice that the index values range from 0% to 100%. 0% means that the muscles of both lower limbs are active at the same percent of the gait cycle (“perfect symmetry”). On the contrary 100% identify a condition of “complete asymmetry”: with reference to the same percent of the gait cycle, the muscle of one lower limb is active while the other one is not.

Table 5.1. Population details. Age, height, weight and the number of analyzed strides for each group are reported as mean \pm standard deviation.

	Number of subjects	Age [years]	Sex	Height [cm]	Weight [kg]	Number of analyzed strides	
CONTROLS	<i>Healthy Children</i>	10	9 \pm 1.5	5M/ 5F	133.3 \pm 7.1	29.1 \pm 5.1	164 \pm 31
	<i>Healthy Adults</i>	10	39.5 \pm 16.2	6M/ 4F	174.4 \pm 9.0	70.4 \pm 13.9	197 \pm 45
	<i>Healthy Elderly</i>	10	69.7 \pm 2.5	5M/ 5F	167.9 \pm 9.1	66.8 \pm 12.1	151 \pm 20
ORTHOPEDIC PATIENTS	<i>Mega TKR</i>	19	37.8 \pm 17.8	10M/ 9F	170.4 \pm 10.5	68.9 \pm 11.4	152 \pm 21
	<i>TKR</i>	10	71.2 \pm 8.8	5M/ 5F	168.2 \pm 9.1	86.8 \pm 21.0	167 \pm 29
	<i>THA</i>	20	66.1 \pm 7.2	11M/ 9F	168.7 \pm 10.5	77.0 \pm 13.3	140 \pm 24
NEUROLOGICAL PATIENTS	<i>Hemiplegic Children</i>	25	8.7 \pm 3.2	15M/ 10F	129.7 \pm 18.8	30.2 \pm 11.7	133 \pm 37
	<i>(iNPH)</i>	10	72.1 \pm 9.6	8M/ 2F	170.8 \pm 7.7	78.3 \pm 10.5	141 \pm 28

5.2.3 Other EMG indices defined in literature

The *EMG_ASYM_INDEX* was then compared with two indices, based on sEMG, that were found in literature. Particularly, the asymmetry index (*ASI*) proposed by Schmidt et al. [68] and the symmetry index (*SI*) proposed by Burnett et al. [69] were computed on the dataset used in this study.

To calculate the *ASI*, the EMG signals, of both lower limbs, acquired during gait were time-normalized to 100% of gait cycle. EMG data were then analysed using the mean muscle activity during the complete gait cycle. The *ASI* is defined as follows (5.2):

$$ASI(\%) = \left| \frac{2 \times (MA_L - MA_R)}{MA_L + MA_R} \right| \times 100\% \quad (5.2)$$

where MA_L and MA_R represent the mean muscle activities of the left and right limb, respectively.

To calculate the *SI*, the RMS values of EMG signals of both lower limbs were computed, and the index values were then obtained by using (5.3):

$$SI = \left| \frac{RMS_{ND,stance}}{RMS_{D,stance}} \right| \quad (5.3)$$

In [69] $RMS_{ND,stance}$ and $RMS_{D,stance}$ were the root mean square amplitude during the stance phase for the non-dominant (ND) and dominant (D) limb, respectively. In the dataset used in this study, however, no information was available about the dominant side of subjects. Consequently, for controls and iNPH patients, the right side was considered as the dominant, while for the remaining groups the sound side was considered as dominant.

5.2.4 Statistical analysis

The mean values and the standard errors of the *EMG_ASYM_INDEX*, *ASI* and *SI* were computed for each group of subjects detailed in Table I and each muscle.

Each group of patients was then matched with one of the three healthy groups (age-based matching), in order to explore the differences between patients and healthy controls. For example, the results obtained for the hemiplegic children were compared with the results obtain on the control group of healthy children.

Then, the Lilliefors test was performed on each group to assess the normality of the distributions obtained applying the three indices to our dataset. Since the Lilliefors test revealed that some of the distributions were not normal, the Wilcoxon

non-parametric test was chosen to compare groups ($\alpha = 0.05$), considering each muscle separately. More specifically, 1-tailed tests were used for comparing *EMG_ASYM_INDEX* and *ASI* values of patients and controls: in fact, the mean value of these indices for patients were expected to be higher than (or equal to) that of controls. Under the same assumption, 1-tailed tests were also applied to investigate the differences in THA patients during the follow-up (between 3 and 6 months, between 6 and 12 months and between 3 and 12 months). On the other hand, 2-tailed tests were using for comparing the *SI* values of patients and controls. In this case, it was not possible to hypothesize a-priori an effect in one direction, due to definition of the index itself: in patients, *SI* may assume values higher or smaller than that of controls, depending on the pathology and the muscle considered.

5.2.5 Results and discussion

Figure 5.3 shows the mean values and the standard errors of the three indices. Observing Figure 5.3a it can be noticed that, as it was expected, the lowest values of the *EMG_ASYM_INDEX* are found in the control groups. Moreover, patients show different behaviors depending on the muscle and the type of pathology they were affected (orthopedic and neurological).

Table 5.2 reports the Wilcoxon test results performed to assess the inter-group differences.

In the following, the results reported in Figure 5.3 and in Table 5.2 will be discuss, particularly comparing each group of patients with the corresponding age-matched control group. However, some considerations can be made a priori, based on the previous knowledge of the different pathologies and treatments which patients underwent. In fact, both orthopedic and neurologic populations were analysed, in which different levels of asymmetry were expected. More specifically:

- (1) Given that the procedure for the implantation of a megaprosthesis implies a higher degree of bone and muscle sacrifice with respect to that of a conventional one, higher differences between Mega TKR patients and controls with respect to those between TKR patients and controls were expected.

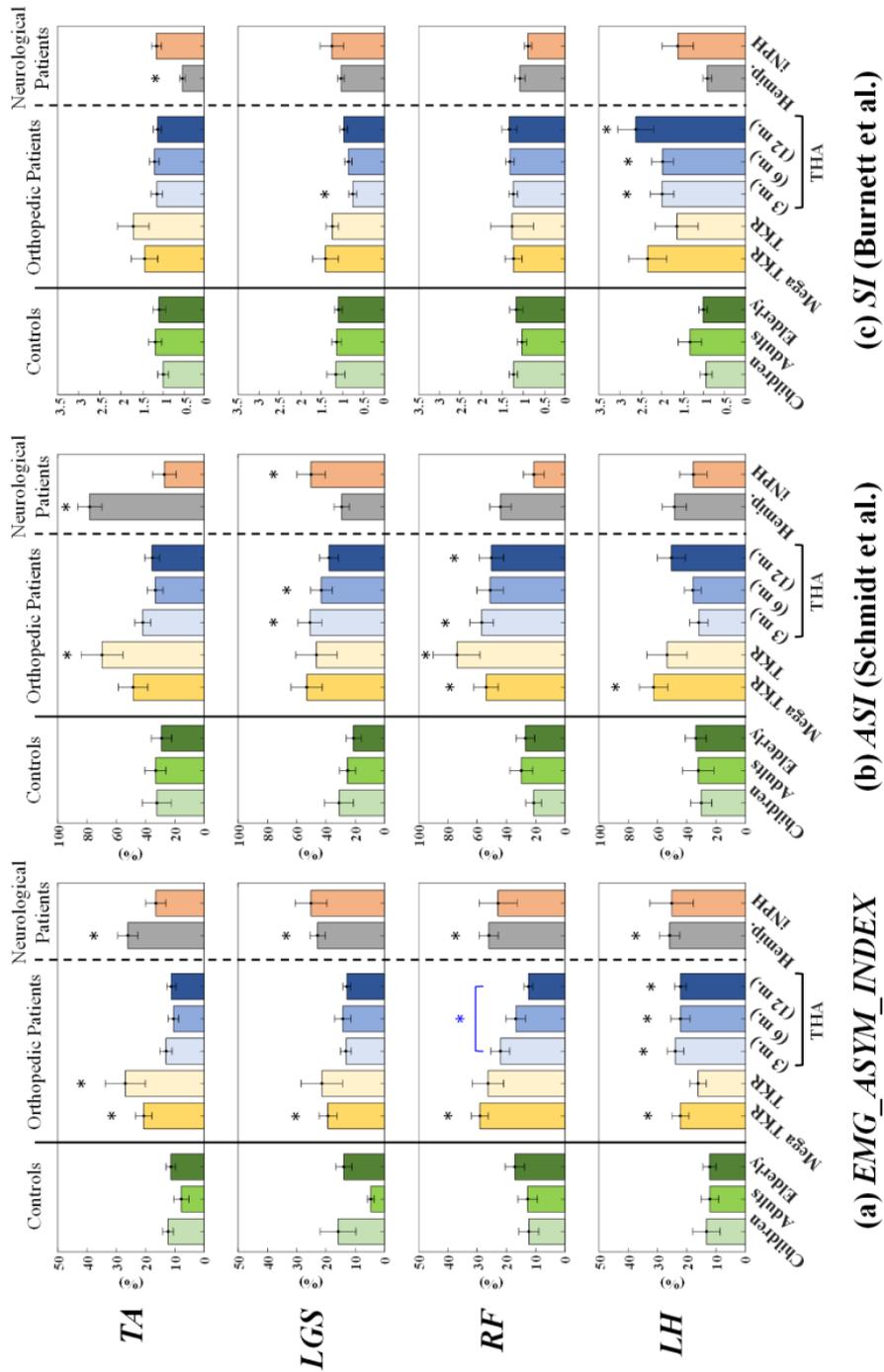


Figure 5.3. Mean values and standard errors of asymmetry indices [70].

“Mean values and standard errors of (a) *EMG_AYM_INDEX*, (b) *ASI* and (c) *SI* on the analyzed groups. Muscle: Tibialis Anterior (TA), Gastrocnemius Lateralis (LGS), Rectus Femoris (RF) and Lateral hamstring (LH). Significant differences between patients and age-matched controls are marked with a black asterisk. Significant difference between THA patients at 3 and 12 months after surgery is marked with a blue asterisk.” [70]

Table 5.2. Comparison between pathological groups and age-matched controls. Wilcoxon-test results.

Group comparison	EMG_ASYM_INDEX <i>p-values</i>						ASI (Schmidt et al.) <i>p-values</i>						SI (Burnett et al.) <i>p-values</i>					
	TA	LGS	RF	LH	TA	LGS	RF	LH	TA	LGS	RF	LH	TA	LGS	RF	LH		
<i>Mega TKR</i>	0.004	0.002	0.002	0.03	0.2	0.09	0.04	0.03	0.9	0.7	1.0	0.3	0.9	0.7	1.0	0.3		
<i>TKR</i>	0.03	0.3	0.1	0.2	0.02	0.08	0.02	0.2	0.6	0.3	0.2	0.5	0.6	0.3	0.2	0.5		
<i>THA 3 months</i>	0.5	0.6	0.2	0.009	0.1	0.02	0.01	0.7	1.0	0.01	0.4	0.02	1.0	0.01	0.4	0.02		
<i>THA 6 months</i>	0.8	0.6	0.6	0.01	0.3	0.02	0.07	0.5	0.8	0.1	0.3	0.004	0.8	0.1	0.3	0.004		
<i>THA 12 months</i>	0.5	0.6	0.8	0.002	0.2	0.06	0.04	0.2	0.8	0.4	0.7	0.003	0.8	0.4	0.7	0.003		
<i>THA 3 months</i>	0.2	0.5	0.08	0.2	0.1	0.2	0.2	0.8	0.4	0.4	0.4	0.9	0.4	0.4	0.4	0.9		
<i>THA 6 months</i>	0.7	0.5	0.3	0.8	0.7	0.4	0.5	0.8	0.9	0.5	0.3	0.4	0.9	0.5	0.3	0.4		
<i>THA 3 months</i>	0.3	0.5	0.01	0.3	0.2	0.1	0.2	0.9	0.7	0.06	0.8	0.4	0.7	0.06	0.8	0.4		
<i>Hemi-plegic iNPH</i>	0.005	0.01	0.004	0.02	0.002	0.4	0.07	0.2	0.02	0.5	0.07	0.9	0.02	0.5	0.07	0.9		
<i>Healthy Children</i>	0.08	0.1	0.4	0.1	0.6	0.01	0.7	0.5	0.3	0.9	0.5	0.3	0.3	0.9	0.5	0.3		
<i>Healthy Elderly</i>	0.08	0.1	0.4	0.1	0.6	0.01	0.7	0.5	0.3	0.9	0.5	0.3	0.3	0.9	0.5	0.3		

*Statistically significant differences (p<0.05) are highlighted in bold font.

(2) The analysis of activation patterns of THA patients reported in the previous chapter (Figure 4.7) did not evidence substantial qualitative differences between prosthetic and sound sides, except for the LH muscle. Hence, for these patients, a limited asymmetry level was expected to be for TA, LGS and RF muscle, while for LH muscle higher value of *EMG_ASYM_INDEX* may be found.

(3) Regarding the two groups of neurological patients, different levels of asymmetry were expected to be found, since hemiplegia is a condition that affects specifically one side of the body, whereas normal pressure hydrocephalus is not known to selectively affect a specific side.

Mega TKR patients and Healthy Adults

The subjects included in the Mega TKR group are patients affected by malignant tumours of the distal part of the femur, who underwent the surgical procedure of modular knee prosthesis (megaprosthesis) implant for saving the lower limb [71], [72]. It is widely known that this kind of surgical procedure implies significant degrees of bone and muscle sacrifice with consequent changes in the gait characteristics of the prosthetic with respect to the sound side [73]. Observing the results in Figure 5.3a and the p-values reported in Table 5.2, it is evident that the *EMG_ASYM_INDEX* reflects these aspects clearly: the index values are lower in the adult controls with respect to the Mega TKR, for all the muscles. Moreover, these differences resulted always statistically significant, as it emerged from the Wilcoxon test results.

TKR patients and Healthy Elderly

The Total Knee Replacement is a more conservative procedure than the megaprosthesis implant and in this case, the *EMG_ASYM_INDEX* reflects this aspect. Overall, the index values are lower in the elderly controls with respect to the TKR group, for all the muscles, but the only significant difference between the two groups has been found for the TA muscle. These results suggest that TKR patient expose asymmetry levels during walking only slightly higher with respect to healthy subjects of the same age.

THA patients and Healthy Elderly

Observing Figure 5.3 some interesting consideration can be made:

- a) TA and LGS muscles show *EMG_ASYM_INDEX* values similar to the ones of elderly controls, highlighting that the procedure of total hip replacement does not produce alteration (for what concerns gait symmetry) in muscles not involved in hip joint control.
- b) The RF muscle shows an interesting behaviour: even if the differences between patients and controls are not statistically different, a qualitative decreasing trend among the three time points is observed. Moreover, between the time points at 3 and 12 months after surgery, the *EMG_ASYM_INDEX* resulted significantly different, suggesting a progressive recovering of symmetrical muscle activation patterns during walking.
- c) As it was expected, LH is the muscle with the greater value of the *EMG_ASYM_INDEX*: at all the time points, statistical differences between patients and controls were found.

Hemiplegic and Healthy Children

Hemiplegia is a common consequence of cerebral palsy (CP) and causes altered selective motor control, weakness and spasticity [74]. This condition causes alterations in motor functions of both upper and lower limbs, selectively on one side only. The results of *EMG_ASYM_INDEX* reflect clearly this aspect of the pathology, since, for all the muscle of this group, the *EMG_ASYM_INDEX* values are significantly higher with respect to the values of healthy children.

iNPH patients and Healthy Elderly

iNPH is a pathology caused by an excess of cerebrospinal fluid in the cerebral ventricles of the brain. Unlike the hemiplegia, this pathology does not affect a specific side of the body, but the overall walking scheme results altered in iNPH patients [75]. The results obtained for iNPH patients are consistent with the previous consideration: no significant differences between this group and controls were found.

5.2.6 Comparison with indices found in literature

Figure 5.3b and 5.3.c show the results obtained applying the *ASI* and *SI* to the analysed dataset and the corresponding *p*-values of the Wilcoxon tests are reported in Table 5.2.

The first importance consideration about these indices is that they cannot be directly compared to the *EMG_ASYM_INDEX*. In fact, both *ASI* and *SI* are related

to the amplitude of the EMG signal during the whole gait cycle, while *EMG_ASYM_INDEX* was defined using principal activations and so it is based on EMG onset-offset timings at each percent of the gait cycle. Hence, the information provided by *EMG_ASYM_INDEX* concerns the differences in the timing of activation patterns of left and right limbs, rather than the differences in the EMG amplitudes.

Therefore, the information obtained using the *EMG_ASYM_INDEX* may provide useful information in addition to those provided using *ASI* and *SI*.

As an example, considering hemiplegic children, *ASI* or *SI* point out significant differences only for the TA muscle, while using *EMG_ASYM_INDEX* significant differences between hemiplegic and healthy children were found for every muscle. This is reasonable since hemiplegic children are expected to show asymmetric timing patterns in every muscle [3] considering also compensation mechanisms. This suggests that, in this case, *EMG_ASYM_INDEX* is more sensitive in identifying walking asymmetry than the other two indices.

5.2.7 Conclusion

A novel EMG asymmetry index based on PAs was presented in this study and applied to a dataset consisting of healthy subjects and both neurological and orthopaedic patients. Different asymmetry levels were expected to be found on each group, considering the different disorders and treatments which patients underwent. The results confirmed this expectation: the value obtained for the asymmetry index was consistent with the expected asymmetry level of each specific population of patients. This suggests that the *EMG_ASYM_INDEX* can be successfully used in clinics for an objective assessment of the asymmetry of muscle activation patterns during locomotion.

5.3 Muscle Functional Indices

In recent years, the objective assessment of pathologies based on gait data has become a research field of great interest. In literature, several works took advantage of gait parameters to improve the diagnostic process of different pathologies such as Parkinson disease [76]-[78] and cerebral palsy, [79], [80], but only a few works used the information extracted from sEMG signals to this purpose [81], [82].

In this context, the definition of a quantitative and reliable index for measuring the distance of the dynamic muscle function of a pathological subject from that of

a reference healthy population can be extremely useful for diagnosis, assessment of the disease progression, and evaluation of treatment outcomes.

The aim of this section is twofold. First, to present a muscle-specific functional index that quantifies the similarity of the activation pattern of a specific muscle of a subject with that of the corresponding muscle of a healthy population. Secondly, to present a global index to quantify the distance between the functionality of a pool of muscles of a specific subject and the functionality of the same muscles of a reference population.

At first the definition of the proposed indices is provided; then, the effectiveness of the muscle-specific and global indices is assessed by applying these indices to a group of 25 healthy children and to a group of 25 hemiplegic children. The purpose was measuring the distance of specific subjects (both at a muscle-specific and global level) as well as of the entire population from a reference population of 55 healthy children.

5.3.1 Definition of the muscle functionality indices

The definition of the muscle functional indices (muscle-specific and global) requires two phases: the representation of the muscle function of the reference population and the calculation of the indices. Each phase is composed of various steps that are described below (Figure 5.4).

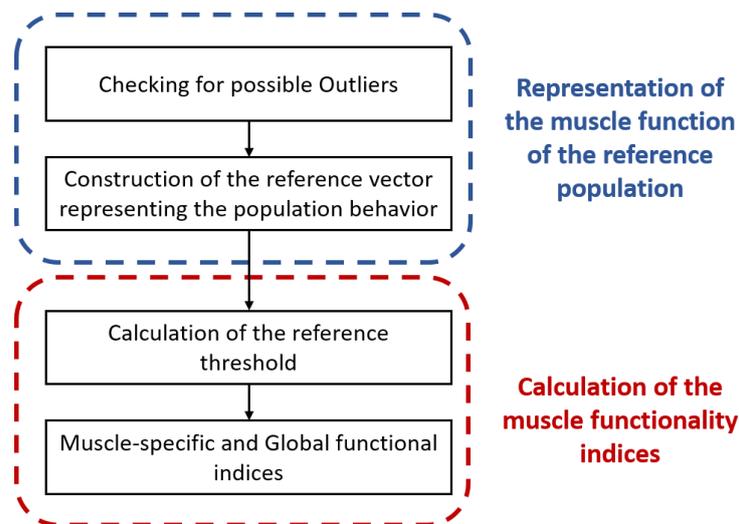


Figure 5.4. Pipeline for the definition of the muscle functional indices.

Representation of the Muscle Function of the Reference Population

At first, the left and right PAs of each muscle are extracted for all the subjects belonging to the reference population using the CIMAP algorithm. Then, the processing for the representation of the muscle function of the reference population consists of 2 steps: a) checking for possible outliers; b) construction of the reference vector representing the population behavior. The two steps are detailed below.

a) Checking for possible outliers

To avoid errors in the representation of the muscle function of the reference population, due to anomalous muscle activations, possible outliers are identified and excluded. The procedure consists in the following steps:

1. Left and right PAs of a specific muscle are grouped together.
2. PAs with the same modality are grouped together.
3. Agglomerative hierarchical clustering is applied to each group (resulting after step 1 and 2), separately. The hierarchical clustering procedure begins with all the PAs considered as single clusters and iteratively merges the two clusters with the lowest distance, until a unique cluster containing all the PAs is obtained. In this specific case, to select the two clusters to be merged at each iteration, the complete linkage is used as linkage method: first, the two farther elements in two clusters are identified, and then the clusters with the lowest distance are merged. The distance between PAs x and y is computed as:

$$D(x, y) = \sum_{i=1}^{1000} |x_i - y_i| + P \quad (5.4)$$

where x_i and y_i are the values of the i -th bit in x and y respectively and P is a penalty term that is set equal to 1000 (that is the maximum difference achievable comparing the two strings) if there is no intersection between corresponding activation intervals (Figure 5.5a) or if the two PAs differ by more than 15% of the gait cycle (Figure 5.5b). The final clusters are obtained by cutting the dendrogram at a specific level: the first iteration (k) in which two penalized clusters are merged together ($D > 1000$) is identified and the dendrogram is cut just before it, considering the clusters obtained at iteration $k-1$.

4. To compensate the unnecessary divisions due to complete linkage, which privileges the minimization of intra-cluster variability [36], a further step is

applied to merge clusters with similar PAs. The centroid of each cluster is defined as the median of the onset-offset timing of PAs belonging to the cluster. Then, the distance $D(x,y)$ in eq. (5.4) between each pair of centroids is iteratively computed and the two clusters having the closest centroids are merged iteratively, until all the distances become greater than 1000.

5. If there are clusters with less than 5% of total number of PAs, they are considered as outliers and hence they are not used for constructing the reference vector.

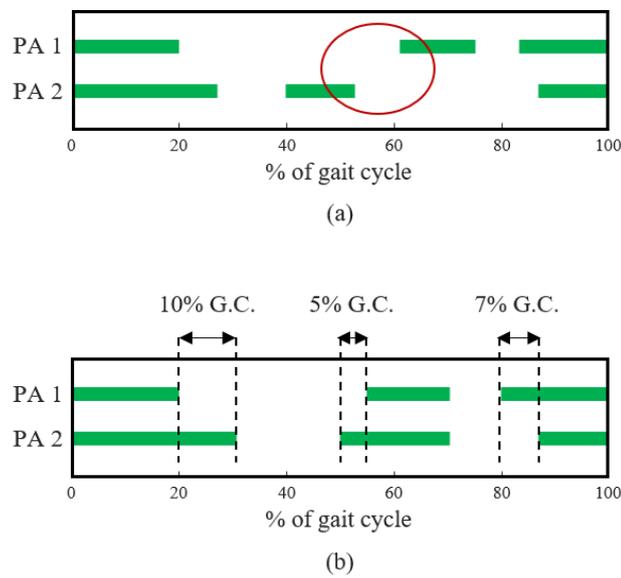


Figure 5.5. Examples of penalty term application.

The penalty term P is equal to 1000 in both the represented situation. In Fig. 5.5a there is no intersection between corresponding activation intervals (see red circle) and in Fig. 5.5b the two PAs differ by 22% of gait cycle (>15%).

Figure 5.6 shows an example of the pipeline for outlier checking, in which outliers are found (Tibialis Anterior muscle, right side).

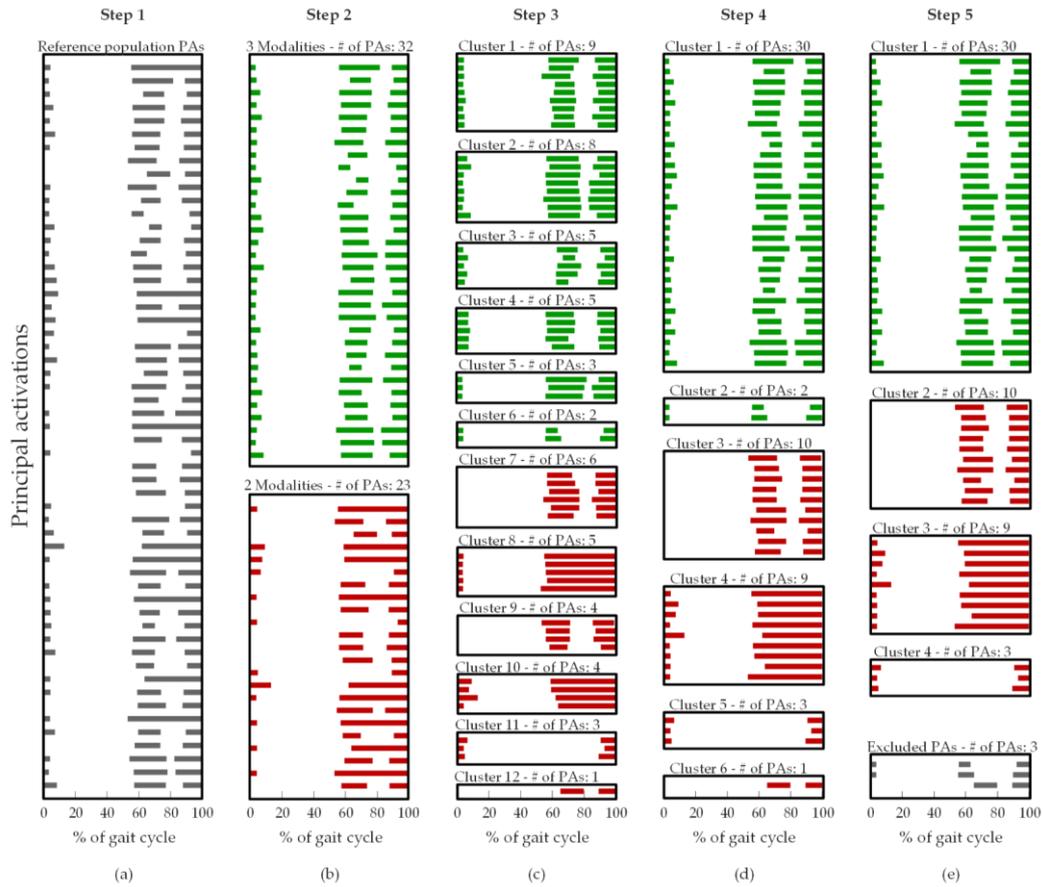


Figure 5.6. Example of the outlier removal pipeline (Tibialis Anterior muscle, right side).

a) Step 1: PAs resulting from CIMAP application; b) Step 2: PAs are grouped into datasets with the same modality; c) Step 3: clusters obtained by dendrogram construction; d) Step 4: clusters resulting by merging similar clusters; e) Step 5: final representative clusters are highlighted in green (3 activations) and red (2 activations); PAs belonging to non-significant clusters (outliers) are represented in grey.

b) Construction of the reference vector representing the population behavior

This step allows describing the behavior of the reference population by aggregating the information contained in the PAs. In particular, since each PA is coded in a binary vector of 1000 binary units, all PAs (both right and left) of a given muscle are pooled together and their bit-by-bit average is computed. In this way, a reference vector (R) of 1000 points in the range $[0, 1]$ is obtained, representing the global behavior of the reference population for a specific muscle (Figure 5.7).

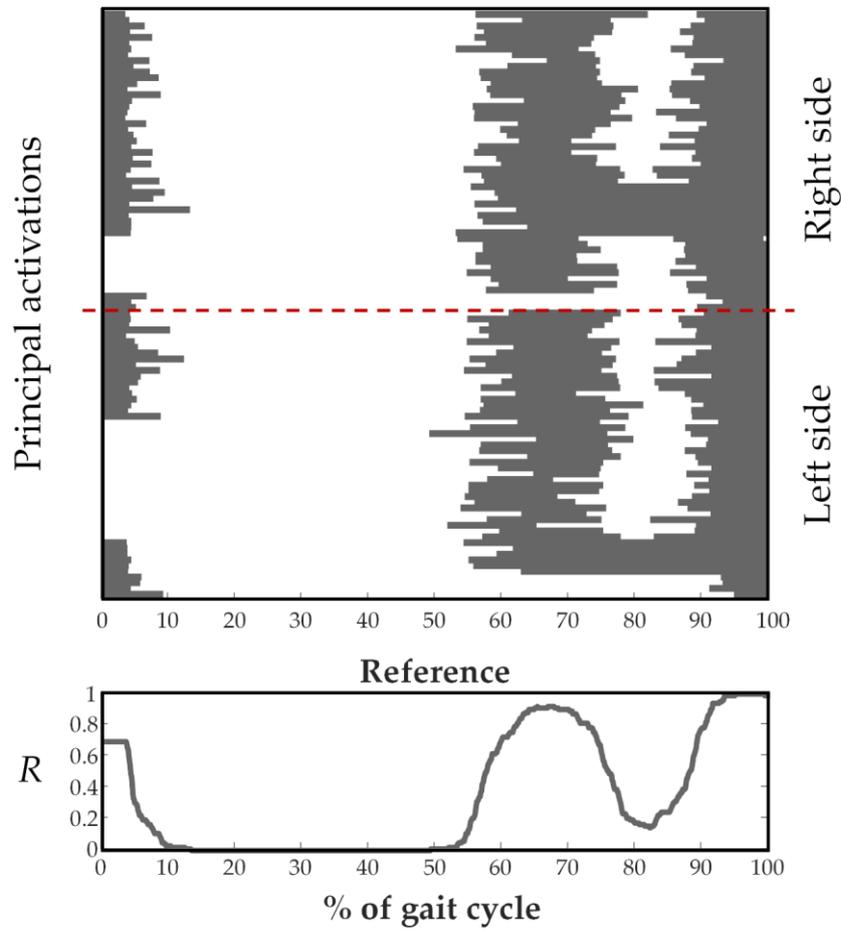


Figure 5.7. Example of the construction and representation of the reference vector R (muscle: Tibialis Anterior).

The reference vector R is obtained by averaging the activation intervals bit by bit across all the principal activations.

Calculation of the Muscle Functional Indices

The calculation of the muscle functional indices consists of two steps, as detailed below. Both steps are based on the measure of the agreement between a specific muscle PA (A) and the reference vector R , evaluated using the Jaccard index (J) [83]:

$$J = \frac{R \cap A}{R \cup A} \quad (5.5)$$

a) Calculation of the reference threshold

The Leave-One-Out (LOO) cross-validation approach is applied to obtain a set of normative J values – one for each muscle – describing the reference population.

For each muscle, the reference vector R is computed using all the subjects but one (R_{LOO_i}). Then, the J value between R_{LOO_i} and the principal activation A_i of the excluded subject is computed using eq. (5.5). This process is carried out for all the subjects belonging to the reference population and a set of J values (one for each muscle of each specific subject) is obtained. The J values are then normalized with respect to the maximum J computed within the population (J_{max}). Then, the probability density function (PDF) of the J value of each muscle is estimated by fitting the J_{norm} distribution using a normal kernel function (Figure 5.8a). Finally, the J_{05} value is computed for each muscle to be used as reference threshold (Figure 5.8b). The J_{05} value is given by eq. 5.6:

$$\int_0^{J_{05}} PDF(J) dJ / \int_0^1 PDF(J) dJ = 0.05 \quad (5.6)$$

It means that the 95% of J values of the reference population lies between J_{05} and 1.

b) Muscle-specific and Global Functional Indices

The muscle-specific functional index J_m is computed according to eq. 5.7:

$$J_m = \frac{J_{A,R}}{J_{max}} \quad (5.7)$$

where $J_{A,R}$ is the Jaccard index calculated between the muscle PA (A) and the corresponding reference vector R , and J_{max} is the maximum J value computed within the reference population.

To obtain the global Muscle Functional Index (MFI) the set of J_m values (one for each muscle) are averaged:

$$MFI = \frac{\sum_{m=1}^M J_m}{M} \quad (5.8)$$

where J_m is the muscle-specific functionality index and M is the number of muscles.

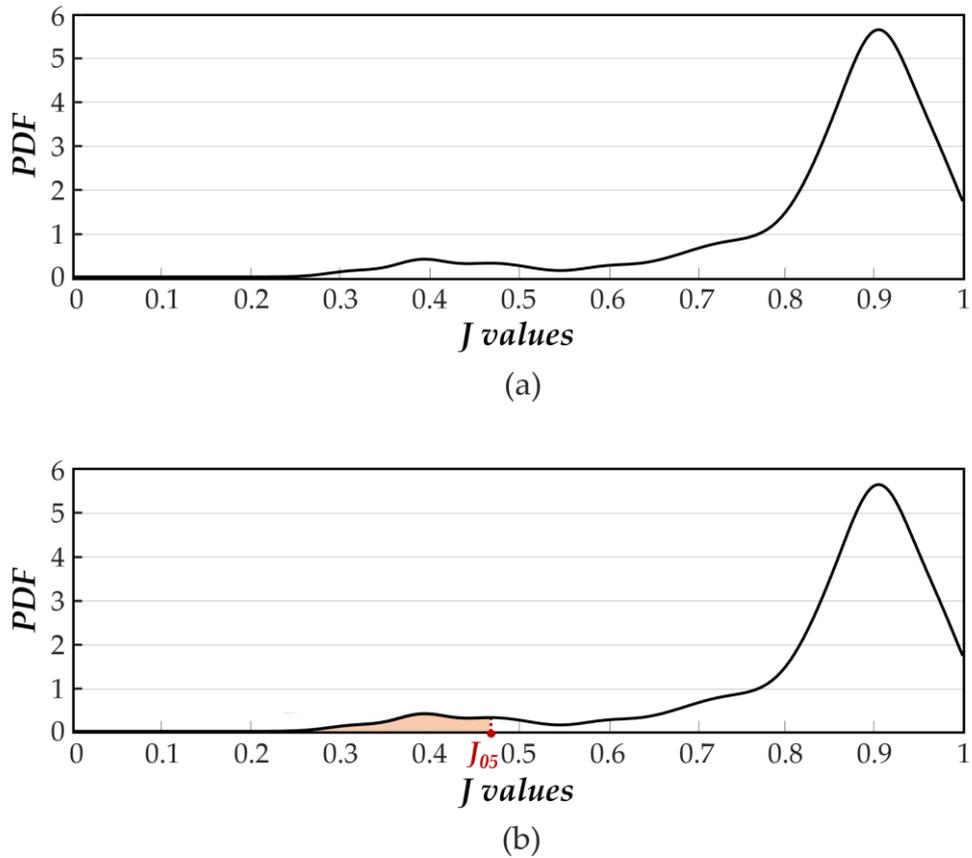


Figure 5.8. Example of probability density function estimated and J_{05} value computing.

a) Example of probability density function (*PDF*) of the J_{norm} distribution (muscle: Tibialis Anterior). b) Example of J_{05} value estimation.

5.3.2 Validation

Population and gait data acquisition

Gait data of 105 school-age children were retrospectively analyzed: 55 typically developing children, without neurological or orthopedic disorders, were used as reference population; 25 healthy children and 25 hemiplegic children were used to evaluate the indices, separately. Table 5.3 reports the details of the populations.

Table 5.3. Population details. Anthropometric characteristics of the sample populations (Healthy and Hemiplegic children). Age, height and weight for each group are reported as mean \pm standard deviation.

	Healthy Children (Reference pop.)	Healthy Children (Test Set)	Hemiplegic Children (Test Set)
<i>Number of subjects</i>	55	25	25
<i>Age (years)</i>	8.9 \pm 1.4	9.0 \pm 1.3	8.7 \pm 3.2
<i>Sex</i>	28M/27F	11M/14F	15M/10F
<i>Height (cm)</i>	133.5 \pm 9.8	132.9 \pm 8.7	129.7 \pm 18.8
<i>Weight (kg)</i>	30.3 \pm 6.2	31.0 \pm 7.4	30.2 \pm 11.7

The muscle activity of Tibialis Anterior (TA), Gastrocnemius Lateralis (LGS), Vastus Medialis (VM), Rectus Femoris (RF) and Lateral Hamstrings (LH) of both lower limbs was analyzed (Figure 5.9).

Subjects walked barefoot at self-selected speed over a 10-m walkway, back and forth for 2.5 minutes.



Figure 5.9. Probes positioning.

Representation of the probe positioning on the analyzed muscles: Tibialis Anterior (TA), Gastrocnemius Lateralis (LGS), Vastus Medialis (VM), Rectus Femoris (RF) and Lateral Hamstrings (LH).

Signal pre-processing

Using the routines included in the STEP32 software, the H, F, P and S phases, of both lower limbs, were obtained. The signal was then segmented in separate gait cycles and time-normalized to the stride duration [13]. For healthy children only the strides showing the normal sequence of gait phases (H, F, P, S) were considered. For hemiplegic children, since a very small number of HFPS strides was available, the strides of the most represented sequence of gait phases of each subject were analyzed [3]. A multivariate statistical filter was used to discard strides related to deceleration, acceleration, and changes of direction [4].

Representation of the Muscle Function of the Reference Population

The three steps described in the previous section were applied to the data of the 55 healthy children to compute the reference vector for each muscle.

Calculation of the Muscle Functional Indices

Equations 5.7 and 5.8 were applied to the data of the 25 healthy children and 25 hemiplegic children to compute the muscle functional indices, both muscle-specific (J_m) and global (MFI). For each muscle, the J_m values are compared with the corresponding reference threshold J_{05} to understand their distance from the reference population.

Statistical Analysis

The Lilliefors test was applied to assess the normality of the MFI distributions of hemiplegic children, both for hemiplegic and sound side, and healthy children, both left and right side. Since the distributions resulted normal, the paired Student t -test ($\alpha = 0.05$) was used to compare: (a) hemiplegic and sound side of hemiplegic children, (b) left and right side of healthy children. The comparisons between healthy and hemiplegic children were performed using the unpaired Student t -test ($\alpha = 0.05$).

5.3.3 Results and discussion

An average of 168 ± 27 gait cycles was collected for each child of the reference population; an average of 167 ± 25 and 133 ± 37 gait cycles were collected for each child of the two test groups (healthy and hemiplegic children, respectively).

In Table 5.4 are reported the results of the Lilliefors test performed among the groups of healthy and hemiplegic children and in Table 5.5 are reported the results of the Student t-tests.

Table 5.4. Lilliefors test results. All the distribution resulted normal distribution (p-values>0.05)

Group	p-values	
	Left side	Right side
<i>Healthy children</i>	0.3	0.08
<i>Hemiplegic children</i>	Hemiplegic side	Sound side
	0.3	0.2

Table 5.5. Student t-test results.

Comparison	p-values
<i>Healthy R – Healthy L</i>	0.1
<i>Hemiplegic S – Hemiplegic H</i>	<0.001
<i>Healthy R – Hemiplegic S</i>	<0.001
<i>Healthy L – Hemiplegic S</i>	<0.001
<i>Healthy R – Hemiplegic H</i>	<0.001
<i>Healthy L – Hemiplegic H</i>	<0.001

*p-values < 0.05 are highlighted in bold.

Figure 5.10 reports the reference vector R obtained for each muscle, as well as the representation of gait phases, i.e. stance (comprising H, F and P) and swing (S phase).

For TA muscle, the reference vector R highlights the typical activation pattern after the 50% of gait cycle (just before toe off and during swing phase) continuing up to the 5% of gait cycle (initial stance), in accordance with [14] and [84]. Moreover, it is evident that the activity during the swing phase is split into two activation intervals for over 80% of the reference population.

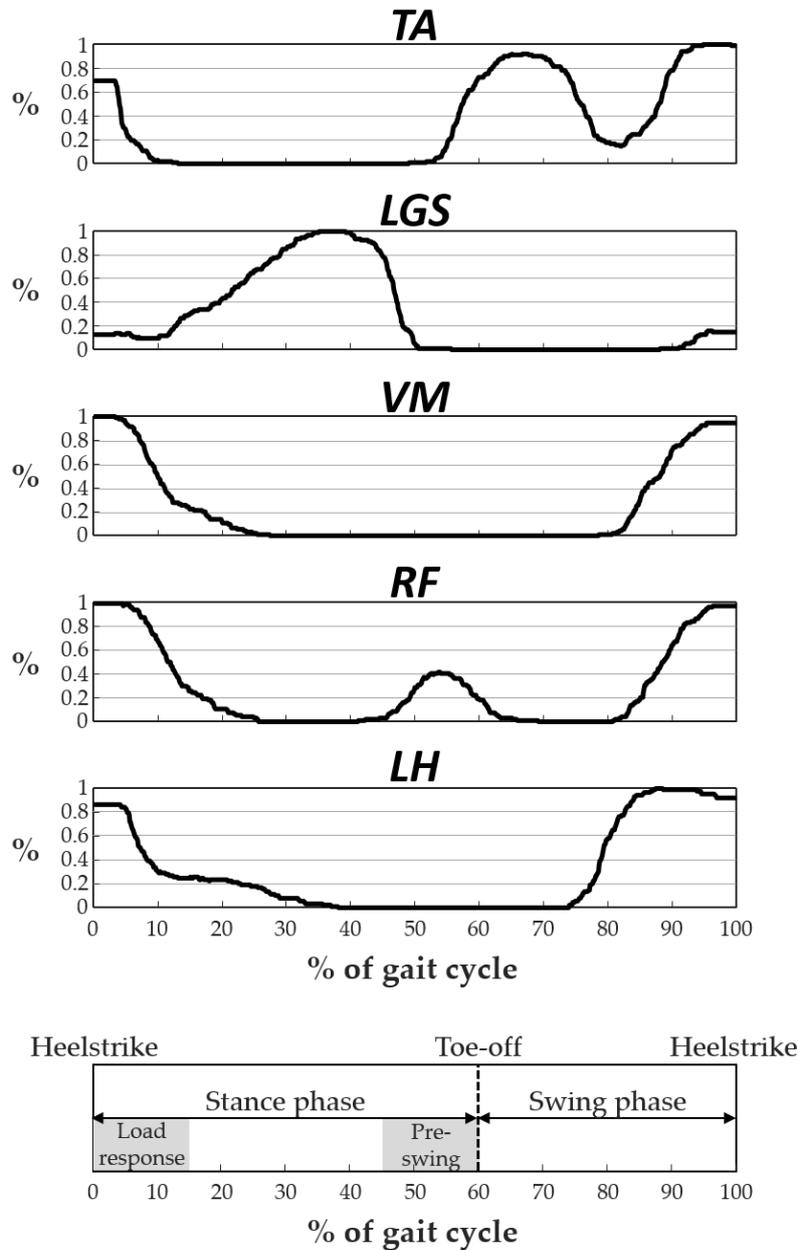


Figure 5.10. Reference vector representation.

Representation of reference vectors R (values expressed as percentage from 0 to 1), for muscles Tibialis Anterior (TA), Gastrocnemius Lateralis (LGS), Vastus Medialis (VM), Rectus Femoris (RF) and Lateral Hamstring (LH). In the bottom plot, a schematic representation of the gait cycle is reported.

For LGS muscle, our results show that most children activate the muscle with a single activation interval that terminates approximately at 50% of the gait cycle. The increasing value of the reference curve between 10% and 35% of gait cycle

points out the difference in the onset of the activation among the subjects belonging to the population.

VM is active from midswing to midstance, as confirmed by [85], who describes a single VM activation pattern. Other studies revealed also an activation in terminal stance (around 50-60% of gait cycle) [14], [86], that is not evident when analyzing PAs. This activation in terminal stance could be aimed at patella stabilization before entering pre-swing phase, but, as mentioned above, this activation is not present in all the strides.

Similarly, the activity of RF extends from midswing to the successive stance phase. However, the RF reference vector highlights a further activation interval, evident in about 40% of the population, before the stance-to-swing transition (45-60% of gait cycle). These results are in agreement with other works found in literature [14], [84].

Finally, the LH reference curve shows a great variability in the activation pattern among the population. Generally, the muscle shows a single activation, lasting from midswing (about 80% of gait cycle) to midstance [14]. However, especially in the midstance phase (between 0 and 30% of gait cycle) it is evident that different types of behavior are present in the population, in terms of duration of the midstance activation.

Figure 5.11 shows the J_m value for each muscle (muscle-specific index) and the final MFI (global index) for two representative subjects (panel a: hemiplegic child; panel b: typically developing child).

The dotted line indicates the reference value J_{05} for each muscle. J_m values above the reference are represented in green, while values below the reference are represented in red. This representation allows highlighting muscles with an abnormal behavior and significantly simplifies the interpretation of the MFI values. Fig. 5.11a shows that, for the hemiplegic child, only the Vastus Medialis of the sound side and the Lateral Hamstring of the hemiplegic side have a J_m value above the reference, meaning that their behavior is similar to that of 95% of the reference healthy population. This situation produces MFI values of 0.40 and 0.38 for the sound and hemiplegic side, respectively. Differently, for the healthy child reported in Fig. 5.11b all the muscles (both sides) present a J_m value above the reference and the final MFI values are equal to 0.90 and 0.85 for left and right side, respectively.

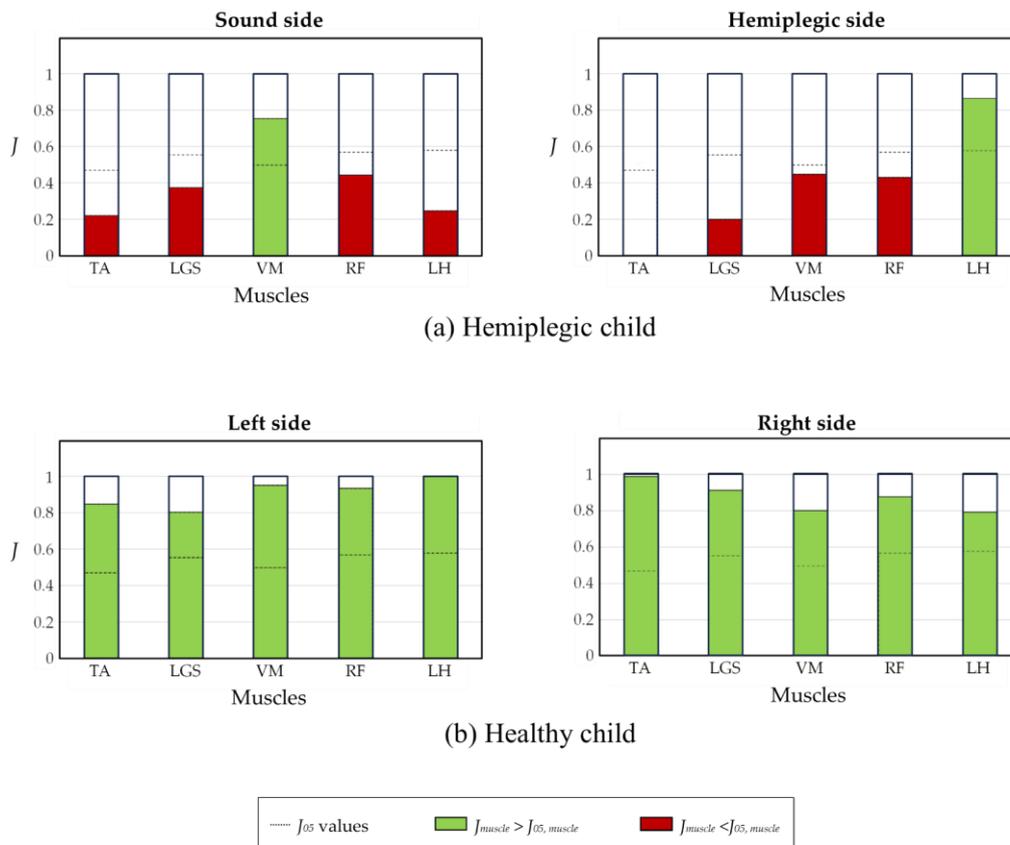


Figure 5.11. Examples of J_m value representation.

Bar diagram representation of the J_m values for (a) a hemiplegic child and (b) a healthy child, both sides. The dotted lines indicate the reference value J_{05} for each muscle. Values lower than J_{05} are represented in red, values higher than J_{05} are highlighted in green and the MFI value is represented in blue. Muscles: Tibialis Anterior (TA), Gastrocnemius Lateralis (LGS), Vastus Medialis (VM), Rectus Femoris (RF) and Lateral hamstring (LH).

Figure 5.12 reports the MFI values (mean value and standard deviation) over the populations of hemiplegic and healthy children (left and right side separately). In particular, the values of the index are not statistically different between the left and right side of healthy subjects ($p = 0.1$), while it is statistically significantly different between each lower limb of healthy children and both the hemiplegic and sound side of hemiplegic children ($p < 0.001$). Finally, also between the hemiplegic and sound sides of hemiplegic children we found a statistically significant difference of the MFI ($p < 0.001$).

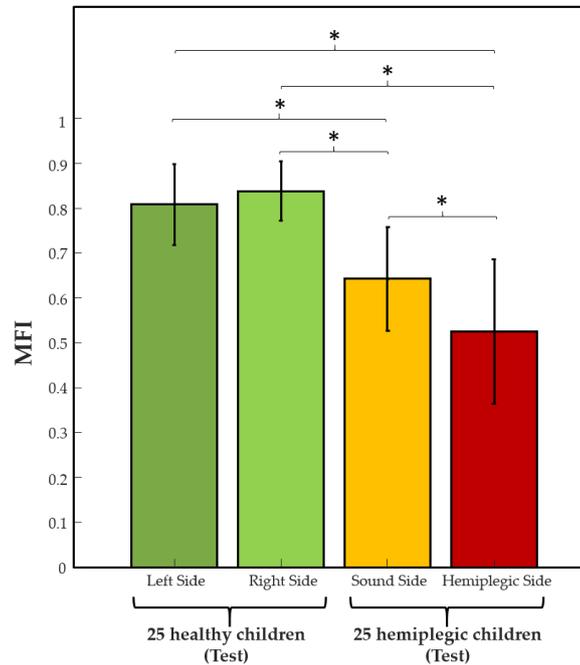


Figure 5.12. *MFI* values (mean value and standard deviation) over the populations of hemiplegic and healthy children.

Mean values (bars) and standard deviation (whiskers) of the *MFI* of healthy and hemiplegic children. Asterisks (*) highlight significant differences between groups ($p < 0.001$).

5.3.4 Conclusion

In this study two muscle functionality indices based on principal activations were introduced. These indices quantify the distance of the functionality of muscles belonging to a specific subject from the average muscle functionality obtained over a reference population. In this way, the overall muscle performance of a specific subject may be quantitatively evaluated for a single muscle (muscle-specific index) and for a specific muscle pool (global index).

When single muscles are considered, the muscle-specific index J_m allows for outlining which muscles of an individual subject show a behavior different from that of the reference population. This information should be important in clinics to allow physicians focusing their attention on those muscles with an abnormal function, also allowing for a quantitative and objective evaluation of abnormality.

MFI and J_m values may be useful in clinics for providing an overall evaluation of muscle functionality during both the first instrumental examination of a subject and when a subject is evaluated successively, along a rehabilitation program.

Chapter 6

Principal and secondary activation extraction applied to muscle synergy analysis

6.1 Introduction

Human locomotion is a complex motor task, due to the multiple degrees of freedom of the skeletal muscle system. Several works in literatures suggested that the central nervous system (CNS) controls a small number of muscle synergies, each one associated to a specific biomechanical function, rather than trying to coordinate the single muscles involved in a motor task [87], [88].

To obtain muscle synergies, the EMG signals of at least 3-15 gait cycles are usually averaged or concatenated [89], [90] and then data reduction techniques, such as Non-negative Matrix Factorization (NMF), are applied to the matrix of EMG signals [91], [92].

As already documented in literature, the extraction of muscle synergies in human locomotion may be influenced by the pre-processing applied to EMG data [92].

The aim of this chapter is to evaluate how the extraction of muscle principal activation may influence the muscle synergy analysis [93]. Then, a pilot study is

reported, evaluating the result obtained extracting muscle synergies using muscle secondary activations [94].

6.2 Gait data acquisition

Gait data acquired from a sample population of 22 healthy subjects (age: 39.2 years \pm 17.0 years, gender: 18 females and 4 males, height: 165.2 cm \pm 8.2 cm, weight: 60.9 kg \pm 17.5 kg) was analyzed. Twelve EMG probes were placed on subject dominant lower limb (twenty subjects out of 22 were right-limb dominant, while two were left-limb dominant) and trunk (Figure 6.1). The probes were positioned over the muscle's belly of: right Longissimus Dorsii (LDR), left Longissimus Dorsii (LDL), Tensor Fasciae Latae (TFL), Gluteus Medius (GMD), Rectus Femoris (RF), Lateral Hamstring (LH), Medial Hamstring (MH), Vastus Medialis (VM), Lateral Gastrocnemius (LGS), Peroneus Longus (PL), Soleus (SOL) and Tibialis Anterior (TA). Foot-switches were used for timing the gait cycle.

All the subjects walked barefoot for 5 minutes at self-selected speed, back and forth on a 10-m straight walkway.

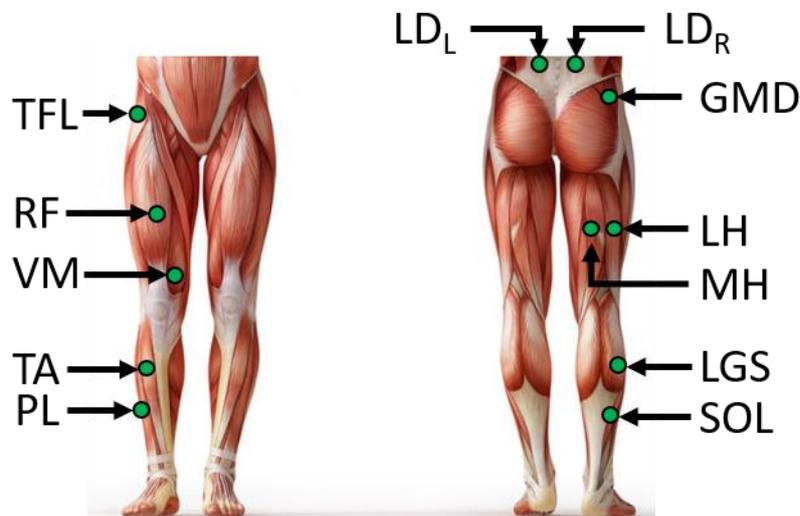


Figure 6.1. Probes positioning.

Representation of the probe positioning on the analyzed muscles: right Longissimus Dorsii (LDR), left Longissimus Dorsii (LDL), Tensor Fasciae Latae (TFL), Gluteus Medius (GMD), Rectus Femoris (RF), Lateral Hamstring (LH), Medial Hamstring (MH), Vastus Medialis (VM), Lateral Gastrocnemius (LGS), Peroneus Longus (PL), Soleus (SOL) and Tibialis Anterior (TA).

6.3 Influence of PA extraction on muscle synergy analysis

Two pre-processing procedures were applied to sEMG data, before muscle synergy extraction: a) the standard approach, in which all the time samples of the sEMG signal were considered and b) a new approach in which only the time samples of the sEMG signal corresponding to principal activations were considered.

6.3.1 Standard pre-processing

In the standard approach, the sEMG signal of each muscle is processed as follows:

- 1) The EMG signal is segmented into separate stride and each stride is time-normalized with respect to gait cycle duration (resampled to 1000 points) [13].
- 2) The signal corresponding to each gait cycle is low-pass filtered at 35 Hz, demeaned, full-cycle rectified and, finally, low-pass filtered at 12 Hz by 5th order Butterworth filter to obtain the sEMG envelope.
- 3) The signals of the various gait cycles are concatenated into a single vector.
- 4) The vector is normalized in amplitude, with respect to the maximum activation
- 5) The vector is divided into N epochs of 10 concatenated gait cycles (epoch 1: gait cycles from 1 to 10, epoch 2: gait cycles from 11 to 20, and so on). The last epoch is discarded if it contains less than 10 gait cycles.

6.3.2 Pre-processing using principal activations

In this kind of pre-processing, the steps are the same as those of the standard one, except that, after the second step, the muscle PAs are extracted using the CIMAP algorithm. The time-normalized EMG signal of each gait cycle are then windowed using a binary mask that is set to 1 in correspondence of muscle principal activation and to 0 if no muscle principal activations is present.

The remaining standard pre-processing steps (step 3, 4 and 5) are applied to the resultant EMG signal.

6.3.3 Muscle synergy extraction

At each subgroup of 10 concatenated gait cycles, the Non-Negative Matrix Factorization (NNMF) algorithm was applied to extract muscle synergies from the filtered sEMG. This algorithm models the original sEMG signals (M) as the linear combination of the time-independent muscle synergy weights (W) and the time dependent activation coefficients (C) [95] as described in eq. 6.1.

$$M(t) = \sum_{k=1}^N C(t)_k \cdot W_k + e \quad (6.1)$$

In eq. 6.1:

- N is optimal number of muscle synergies necessary for describing the motor task;
- W_k is the weight vector and it describes the contribution of each observed muscle to the k -synergy;
- $C(t)_k$ is the activation coefficient vector and it represents the time-dependent modulation of the muscles recruited in the k -synergy;
- e is the prediction error of the factorization algorithm.

More details of the implemented algorithm can be found in [94].

The optimal number of muscle synergies (N_{90}) was then chosen using the total Variance Accounted For ($tVAF$) indicator [94].

Finally, the weight vectors W_k were normalized in amplitude with respect to their global maximum, and the activation coefficient vectors $C(t)_k$ were multiplied by the correspondent normalized values.

6.3.4 Muscle synergy analysis

The muscle synergies extracted using the two approach (standard and PAs based), were compared evaluating the number of muscle synergies and the interpretability of the results [96].

The number of muscle synergies was expressed using the N_{90} value, described before, while muscle synergy interpretability (I) was evaluated considering the average of the muscle synergy weights that are not directly involved in the biomechanical function described by the k -synergy. In fact, a muscle synergy can be considered more easily interpretable when the values of the weights of the muscles not directly involved in the specific biomechanical function are close to zero, while it can be considered less interpretable when they have values comparable with the weights of the muscles directly involved in the biomechanical function. The I values are expressed in percentage, and range between 0% (low interpretability) and 100% (high interpretability).

6.3.5 Results and discussion

On the average, a dataset of 156 ± 25 HFPS gait cycles was analyzed for each subject. In the following, the results obtained with the standard and the novel pre-processing approach will be discussed.

The same number of muscle synergies ($N_{90}=5$ for all the 22 subjects) was needed to properly reconstruct the original sEMG signals for both the approaches. Figure 6.2 reports an example of the muscle synergies extracted from a representative subject of the sample population using the two pre-processing techniques [96].

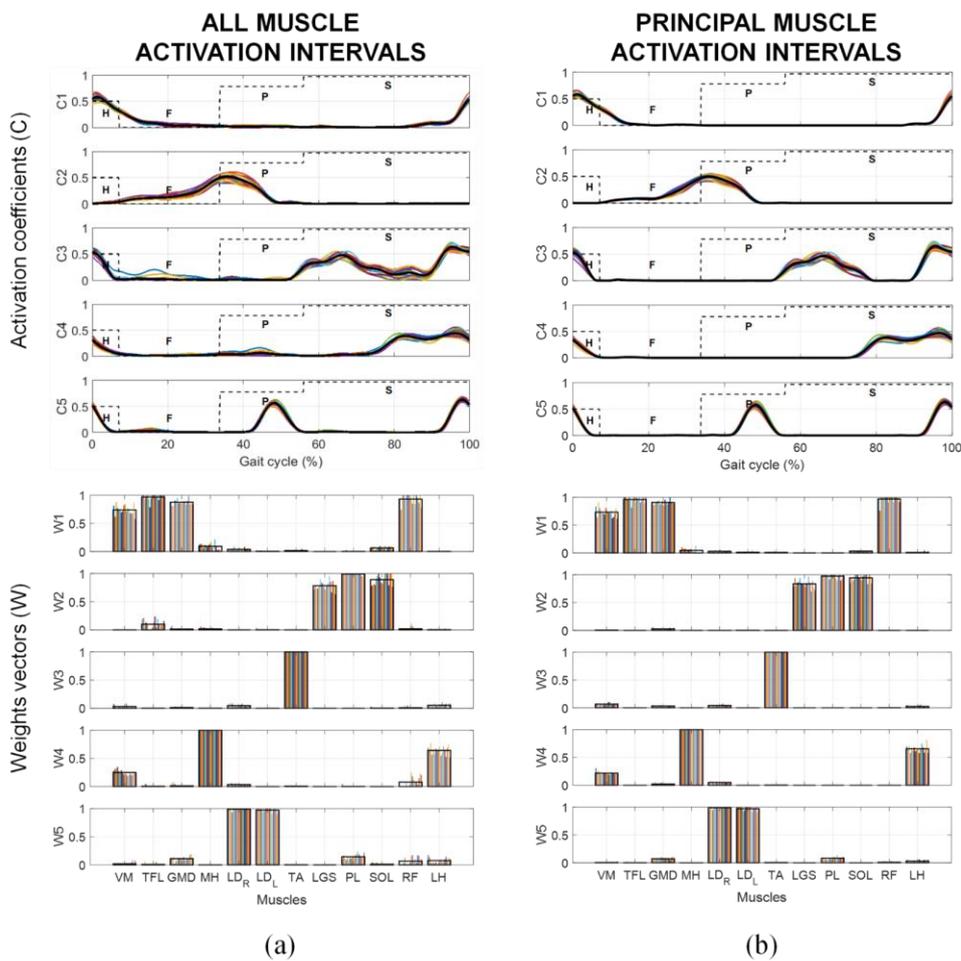


Figure 6.2. Activation coefficients C_k and weight vectors W_k obtained with two different processing techniques. [96]

a) Standard approach, and b) novel approach with extraction of principal activations (PAs). Each colored line (or colored vertical bar) represents C_k (or W_k) extracted from a single subgroup of 10 gait cycles. Black lines (or top of black rectangles) represent the average C_k (or W_k) across subgroups. The dotted lines, in the C_k -plots, represent the mean footswitch signal with the indication of the 4 gait phases: Heel contact (H), Flat foot contact (F), Push off (P) and Swing (S).

On the other hand, the computed I -value was equal to $90.45\% \pm 0.52\%$, for the standard approach, and $92.72\% \pm 0.62\%$, for the novel approach. Moreover, results revealed a statistically significant increase in the interpretability of the muscle synergies extracted from the sEMG signals processed by means of the novel approach with respect to those extracted by applying the standard approach (Student t -test result: $p = 0.0007$).

The muscle synergy analysis is a topic of great relevance and several studies supporting the importance of this theory have been published in the last years [97], [98], [99]. The standard approaches used to extract the muscle synergies generally consider the whole sEMG signals as input of the factorization algorithm [90], [92], [100]. However, the great intra-subject variability of the sEMG activation patterns may significantly influence the results obtained using these approaches [101].

The introduction of muscle PA extraction in the pre-processing pipeline of muscle synergy analysis may result useful to overcome the drawback of the standard techniques. According to this novel approach, only the sEMG time-samples in correspondence of PAs are considered as inputs of the factorization algorithm for the extraction of the muscle synergies while the remaining time-samples are set to zero.

The results obtained comparing the standard and new approach at first reveal that both the approaches needed five muscle synergies to accurately reconstruct the original sEMG signals. Moreover, the extracted muscle synergies resulted very similar both in their composition (profile of the activation coefficients and weighted contribution of the muscles) and in the biomechanical functions produced by each of them. This suggested that no information was lost due to the extraction of the PAs, with respect to the standard approach.

On the other hand, in terms of interpretability of the muscle synergies, the novel approach outperforms the standard one, suggesting that the extraction of the PAs allows for obtaining a better interpretability of the muscle synergies with respect to the standard approach.

The approach based on PAs extraction was validated considering a population of healthy subjects, but the use of PAs for extracting muscle synergies may be a promising technique for analyzing subjects affected by neurological disorders, for whom the assessment of motor control through muscle synergies may be of the utmost importance [102], [103].

6.4 Influence of SA extraction on muscle synergy analysis

In Chapters 4 and 5 the concept of *principal activation* was introduced and widely discussed, and it was anticipated that this concept is complementary to the concept of *secondary activation* (SA). Particularly, secondary activations were defined as those activations present only in some strides: they have an auxiliary function in motor control and are expected to be possibly related to abnormalities or peculiarities of specific subjects.

In practice, starting from the CIMAP clustering results of a specific muscle (Figure 6.3a), each significant prototype may produce secondary activations: they are the obtained as the difference between the binary string coding the prototype and the binary string coding the principal activation (Figure 6.3b).

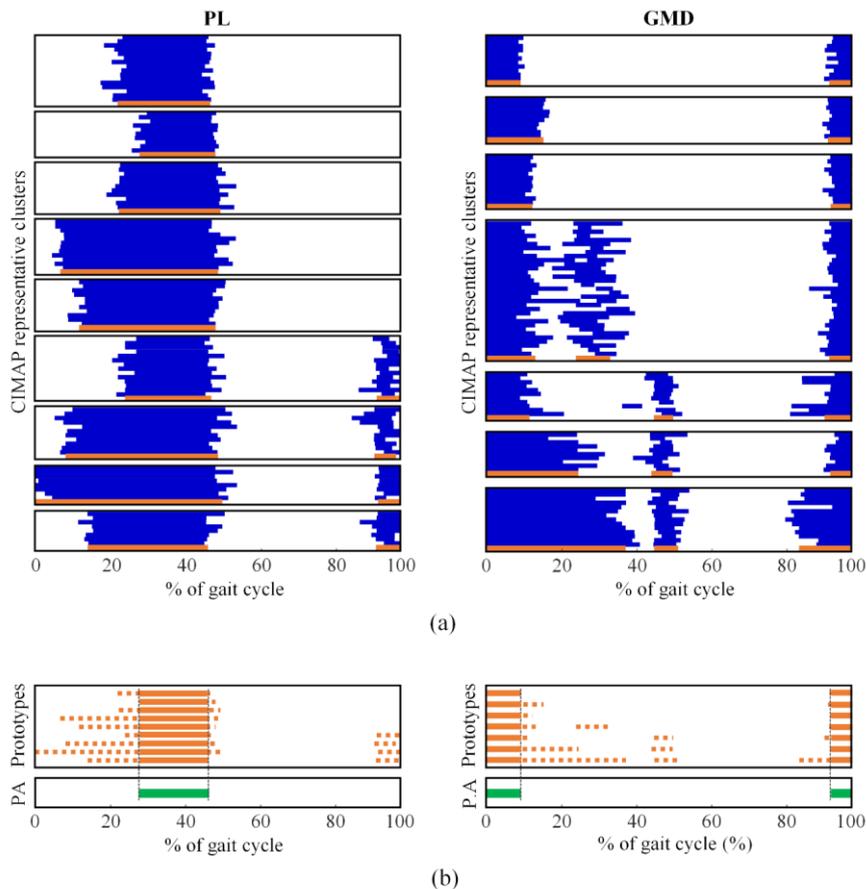


Figure 6.3. Example of secondary activation extraction.

(a) Blue intervals represent the cluster elements (sEMG activation intervals) normalized with respect to the gait cycle duration for PL (left) and GMD (right) muscles, while orange intervals represent the cluster prototypes. (b) Principal activations (PA) are highlighted in green and are defined as the intersection of all the clusters' prototypes extracted through CIMAP. The secondary activations resulting from each prototype are highlighted using a dotted line.

In this pilot study, the sEMG signals acquired from one healthy female volunteer (age: 25 years, height: 160 cm and weight: 61 kg) were analyzed.

6.4.1 Pre-processing using secondary activations

At first, the secondary activations defined as *completely auxiliary* are identified: they are defined as those SAs that are not an extension of a PA interval. Then a binary mask (1000 bit) obtained as the union of all the *completely auxiliary SAs* is defined (Figure 6.4). The binary mask is set equal to 1 in correspondence of at least one muscle *completely auxiliary SA* and to 0 if no muscle *completely auxiliary SA* is present. The time-normalized EMG signal of each gait cycle are then windowed using the binary mask of *completely auxiliary SAs*.

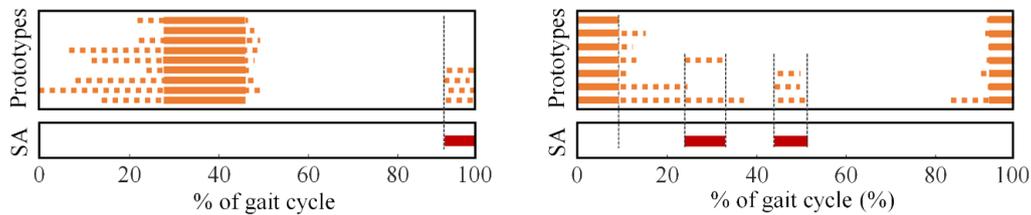


Figure 6.4. Example of *completely auxiliary* secondary activation extraction.

Orange intervals represent the cluster prototypes. The secondary activations resulting from each prototype are highlighted using a dotted line. Red intervals represent the *completely auxiliary* secondary activation.

6.4.2 Results and discussion

Figure 6.5 reports the muscle synergies extracted from the analyzed subject using the standard pre-processing approach (figure 6.5a), the pre-processing applying PA extraction (Figure 6.5b) and the pre-processing applying SA extraction (Figure 6.5c) [94].

It can be noticed that the C_k and W_k computed using the secondary activation intervals are significantly different with respect to the ones represented in Figure 6.5a and 6.5b in terms of number of synergies, shapes and biomechanical function.

The activation coefficients represented in Figure 6.5c result characterized by a lower amplitude and a significantly higher dispersion of the neural commands among the subgroups with respect to the activation coefficients computed using the first two pre-processing procedures. Furthermore, it can be noted that these synergies are actually associated to single-muscle contributions, so rather than highlight muscle synergies obtained processing only the sEMG signal corresponding to secondary activations, this type of analysis allows to highlight when and which muscle contribute in those slight corrections to which secondary activations are attributed.

In fact, it can be observed that the muscular contribution in this case appears at 0, 50 and 100% of gait cycle. In this timing there is the load passage from one lower limb to the other and therefore some small corrections in muscle activation may take place to maintain balance and properly control the trunk and joints of the hip, ankle and knee.

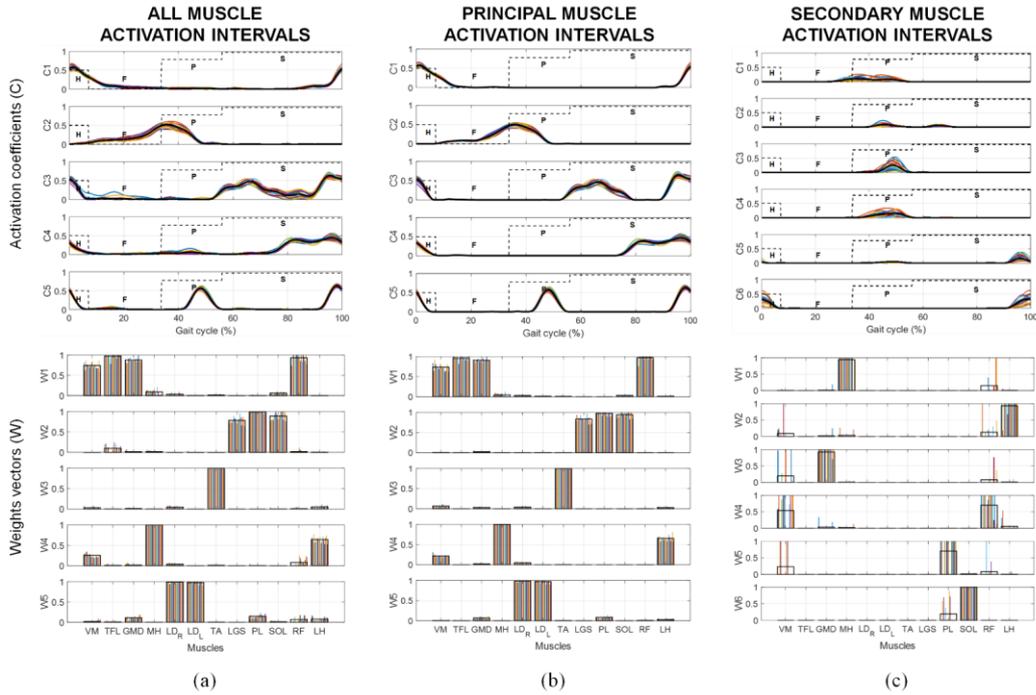


Figure 6.5 Activation coefficients C_k and weight vectors W_k obtained with three different processing techniques. [94]

a) Standard approach, b) approach with extraction of principal activations (PAs) and c) approach with extraction of secondary activations (SAs). Each colored line (or colored vertical bar) represents C_k (or W_k) extracted from a single subgroup of 10 gait cycles. Black lines (or top of black rectangles) represent the average C_k (or W_k) across subgroups. The dotted lines, in the C_k -plots, represent the mean footswitch signal with the indication of the 4 gait phases: Heel contact (H), Flat foot contact (F), Push off (P) and Swing (S).

6.4.3 Conclusion

In the presented study, the impact of the selection of principal and secondary muscle activation intervals from sEMG signals on the muscle synergies extracted during a walking task in healthy subjects was analyzed.

The results obtained using PAs suggested that this kind of pre-processing provide more consistent and more stable activation coefficients and weights vectors,

providing a more interpretable assessment of the modular organization of the CNS during a walking task without any loss of information.

Moreover, the extraction of secondary activation has proved to be an interesting pre-processing step to integrate the information obtained using PAs, in a clear and compact way.

Further step would be the application of this novel approach on sEMG signals acquired from subjects with musculoskeletal or neurological disorders (e.g. Parkinson's disease) during gait to assess its applicability also in pathological conditions. re consistent and more stable activation coefficients and weights vectors.

Chapter 7

Conclusions

The purpose of this thesis was to present new tools and methodologies for analysing muscle activation characteristics during gait.

First of all, a new method (CIMAP – Clustering for Identification of Muscle Activation Patterns) was developed to cope with the problem of the large intra-subject variability in muscle activation onsets and offsets during gait. The CIMAP method is based on agglomerative hierarchical clustering and it allows for grouping strides in clusters presenting similar sEMG activation patterns. In the first part of the thesis the method was presented and tested on healthy subjects. Then, some issues emerged when its application was extended to pathological subjects; an in-depth analysis of the method was performed, and it was optimized to obtain a method usable both in healthy and pathological conditions. The results obtained applying the CIMAP method to groups of healthy subjects and orthopaedic and neurological patients have proved that this method definitively improves the correct interpretation of the EMG signals. At first the method provides an organized representation of the most common type of muscle activation patterns of the individual, that may be useful to clinicians to simply identify possible criticism in subject walking. Moreover, as a result of the CIMAP, cluster centroids are obtained: they are elements that characterize all the strides grouped in a specific cluster. In this way, the stride variability can be represented using only a few elements, representing the main activation patterns exposed during gait by the analysed subject.

Then, the concept of *principal activation* was introduced. From the biomechanical point of view, the principal activations are those activation necessary for ac-

complishing a specific motor task (in this case, walking) and they describe the essential contributions of a specific muscle to the movement. In practice, the principal activations of each muscle were obtained as the intersection of the corresponding cluster centroids, obtained using CIMAP. The concept of principal activations is complementary to the concept of *secondary activations*, which are activations present only in some strides and have an auxiliary function in motor control (e.g. to provide a slight correction to muscle activations due to temporary subject distractions or extemporaneous external disturbances).

Thus, as the cluster centroids allowed to represent in a compact way the subject stride-to-stride variability, similarly the principal activations allowed to characterize the subject through a single activation pattern for each analysed muscle. These patterns are able to provide a synthetic representation of those muscular activations strictly necessary to perform the gait task.

The aggregation of the results obtained applying the CIMAP method and extracting the principal activations, allowed for performing complex analysis of entire group of subjects, sharing homogeneous characteristics. More specifically, three studies that used principal activations to characterize groups of subjects were reported.

In the first study, a population of 100 children school-age children was characterized using principal activations only. The results of this study showed that, considering principal activations only, it was possible to obtain an easy representation of the biomechanical contribution of the observed muscles, very useful for the clinicians. Moreover, it was possible to identify and describe phenomena related to gait maturation that were less evident when all the muscle activations were considered.

The second study dealt with the problem of comparing datasets from a patient cohort with respect to a control group. Particularly, the patient cohort consisted of subjects affected by idiopathic normal pressure hydrocephalus (iNPH) and the control group was a group of age-matched healthy subjects. The analysis of muscle principal activations was applied to both the datasets and it has proved to be a powerful tool to compare muscle activation patterns of the two different groups. Moreover, the concept of gait phase time normalization was introduced in this study and it has revealed to be a valid approach to extract information, especially when significant differences are present in gait phase duration of the group of interest. This study was only an example of the application of principal activation analysis aimed to the comparison of two datasets: this approach can be easily extended for the analysis of other pathologies, becoming a promising tool for clinicians to identify the main differences between healthy and pathological.

Finally, the third study was a longitudinal assessment of muscle function in patients that underwent a total hip arthroplasty surgery. Gait data acquired from 20 patients at 3, 6 and 12 months after surgery and 20 age-matched controls were analysed, in order to assess the efficacy of principal activation analysis in a longitudinal study. The analysis of principal activations allowed for comparing, in a compact way, the results from a longitudinal study conducted on patients after hip replacement surgery and, as in the previous study, the methodology can be easily extended to the analysis of other longitudinal studies.

In the second part of the thesis, two studies were presented aimed to introduce instruments, based on muscle principal activations, for the quantification and interpretation of specific aspects of the gait.

In the first study, the *EMG_ASYM_INDEX* was introduced. This index can assess in an objective and repeatable way the muscle-activation asymmetry in cyclic movements. The index definition was based on muscle principal activations and it was applied to a dataset consisting of healthy subjects, neurological patients and orthopaedic patients. Based on the knowledge of the different disorders and treatments which patients underwent, different asymmetry levels on each population were expected to be found. The *EMG_ASYM_INDEX* values obtained on the different groups were consistent with the expected asymmetry level of each specific population of patients. This suggests that the *EMG_ASYM_INDEX* can be successfully used in clinics for an objective assessment of the asymmetry of muscle activation patterns during locomotion.

In the second study, two other indices are presented. The first one (J_m) is a muscle-specific functionality index that quantifies the similarity of the activation pattern of a specific muscle of a subject with that of the corresponding muscle of a healthy population. The second one is a global index (*MFI*) to quantify the distance between the functionality of a pool of muscles of a specific subject and that of a reference population. Using both these indices, the overall muscle performance of a specific subject may be quantitatively evaluated for a single muscle (J_m) and for a specific muscle pool (*MFI*). *MFI* and J_m indices revealed to be promising useful instruments in clinics for providing an overall evaluation of muscle functionality during both the first instrumental examination of a subject and when a subject is evaluated successively, along a rehabilitation program.

Finally, a brief introduction to the application of principal and secondary activation extraction to muscle synergy analysis was provided. The presented study showed the impact of the selection of principal and secondary muscle activation intervals from sEMG signals on the robustness of the muscle synergies extracted

during a walking task in healthy subjects. The results obtained using principal activation suggested that this kind of processing provide a more interpretable assessment of the modular organization of the central nervous system during a walking task without any loss of information. Moreover, the extraction of secondary activation has proved to be an interesting pre-processing step to integrate the information obtained using principal activations, in a clear and compact way. The further step of this application would be the analysis of patients with musculoskeletal or neurological disorders (e.g. Parkinson's disease), to assess its applicability also in pathological conditions.

In conclusion, in this thesis, several instruments have been developed and validated to analyze the muscle activation characteristics during gait. These tools have revealed to be powerful tools for the simplification of gait data interpretation and for the quantification of gait characteristics.

These tools are not used for diagnostic purposes, but their aim is to support research and care delivery in patients with pathologies that affect locomotion. For research, they can be used to characterize abnormalities of muscle activation patterns in a group of subjects or to enhance the results of other algorithms. For patient care delivery they can be used to decide about physical therapy, to assess the results of a rehabilitation protocol or a surgical intervention.

References

- [1] Perry, J. Gait analysis: normal and pathological function. Charles B. Slack, Thorofare, NJ 1992.
- [2] Agostini, V.; Knaflitz, M.; Nascimbeni, A.; Gaffuri, A. Gait measurements in hemiplegic children: An automatic analysis of foot-floor contact sequences and electromyographic patterns. In Proceedings of the 2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA); IEEE, 2014; pp. 1–4.
- [3] Agostini, V.; Nascimbeni, A.; Gaffuri, A.; Knaflitz, M. Multiple gait patterns within the same Winters class in children with hemiplegic cerebral palsy. *Clin. Biomech.* 2015, 30, 908–914.
- [4] Agostini, V.; Ganio, D.; Facchin, K.; Cane, L.; Moreira Carneiro, S.; Knaflitz, M. Gait parameters and muscle activation patterns at 3, 6 and 12 months after total hip arthroplasty. *J. Arthroplasty* 2014, 29, 1265–1272.
- [5] Agostini, V.; Lanotte, M.; Carlone, M.; Campagnoli, M.; Azzolin, I.; Scarafia, R.; Massazza, G.; Knaflitz, M. Instrumented Gait Analysis for an Objective Pre-/Postassessment of Tap Test in Normal Pressure Hydrocephalus. *Arch. Phys. Med. Rehabil.* **2015**, 96, 1235–1241.
- [6] Chambers, H.G.; Sutherland, D.H. A practical guide to gait analysis. *J. Am. Acad. Orthop. Surg.* **2002**, 10, 222–231.
- [7] Al-Zahrani, K.S.; Bakheit, M.O. A historical review of gait analysis. *Neurosciences (Riyadh)*. **2008**, 13, 105–8.
- [8] Agostini, V.; Knaflitz, M. “Statistical gait analysis”, in *Distributed diagnosis and home healthcare (D₂H₂)*; 2012; ISBN 1588832376.
- [9] Shiavi, R.; Freeman, F.; Griffin, P. *Variability of Electromyographic Patterns for Level-Surface Walking through a Range of Self-Selected Speeds α* ; 1981; Vol. 18;.
- [10] Winter, D.A.; Yack, H.J. EMG profiles during normal human walking: stride-to-stride and inter-subject variability. *Electroencephalogr. Clin. Neurophysiol.* **1987**, 67, 402–411.
- [11] Yang, J.; Winter, D. Electromyographic amplitude normalization methods. *Arch. Phys. Med. Rehabil.* 1984, 65, 517–521.
- [12] Arsenault, A.B.; Winter, D.A.; Marteniuk, R.G. Is there a “normal” profile of EMG activity in gait? *Med. Biol. Eng. Comput.* **1986**, 24, 337–343.
- [13] Agostini, V.; Balestra, G.; Knaflitz, M. Segmentation and Classification of Gait Cycles. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2014**, 22, 946–952.
- [14] Agostini, V.; Nascimbeni, A.; Gaffuri, A.; Imazio, P.; Benedetti, M.G.; Knaflitz, M. Normative EMG activation patterns of school-age children during gait. *Gait Posture* **2010**, 32, 285–9.
- [15] Agostini, V.; Lo Fermo, F.; Massazza, G.; Knaflitz, M. Does texting while walking really affect gait in young adults? *J. Neuroeng. Rehabil.* **2015**, 12, 1–10.

- [16] Di Nardo, F.; Fioretti, S. Statistical analysis of surface electromyographic signal for the assessment of rectus femoris modalities of activation during gait. *J. Electromyogr. Kinesiol.* **2013**, *23*, 56–61.
- [17] Severini, G.; Conforto, S.; Schmid, M.; D'Alessio, T. Novel formulation of a double threshold algorithm for the estimation of muscle activation intervals designed for variable SNR environments. *J. Electromyogr. Kinesiol.* **2012**, *22*, 878–885.
- [18] Severini, G.; Conforto, S.; De Marchis, C.; Schmid, M.; D'Alessio, T. A SNR-independent formulation of a double threshold algorithm for the estimation of muscle activation intervals. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS* **2011**, 7500–7503.
- [19] Vannozzi, G.; Conforto, S.; D'Alessio, T. Automatic detection of surface EMG activation timing using a wavelet transform based method. *J. Electromyogr. Kinesiol.* **2010**, *20*, 767–772.
- [20] Xu, Q.; Quan, Y.; Yang, L.; He, J. An adaptive algorithm for the determination of the onset and offset of muscle contraction by EMG signal processing. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2013**, *21*, 65–73.
- [21] Agostini, V.; Balestra, G.; Knaflitz, M. Segmentation and classification of gait cycles. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2014**, *22*, 946–952.
- [22] Bonato, P.; D'Alessio, T.; Knaflitz, M. A statistical method for the measurement of muscle activation intervals from surface myoelectric signal during gait. *IEEE Trans. Biomed. Eng.* **1998**.
- [23] Rosati, S.; Agostini, V.; Balestra, G.; Knaflitz, M. Basographic gait impairment score: A fuzzy classifier based on foot-floor contact parameters. *IEEE MeMeA 2014 - IEEE Int. Symp. Med. Meas. Appl. Proc.* **2014**.
- [24] Strazza, A.; Mengarelli, A.; Fioretti, S.; Burattini, L.; Agostini, V.; Knaflitz, M.; Di Nardo, F. Surface-EMG analysis for the quantification of thigh muscle dynamic co-contractions during normal gait. *Gait Posture* **2017**, *51*, 228–233.
- [25] Duda, R.O.; Hart, P.E. (Peter E.; Stork, D.G. *Pattern classification*; ISBN 111858600X.
- [26] Den Otter, A.R.; Geurts, A.C.H.; Mulder, T.; Duysens, J. Abnormalities in the temporal patterning of lower extremity muscle activity in hemiparetic gait. *Gait Posture* **2007**, *25*, 342–352.
- [27] Gligorijević, I.; van Dijk, J.P.; Mijović, B.; Van Huffel, S.; Blok, J.H.; De Vos, M. A new and fast approach towards sEMG decomposition. *Med. Biol. Eng. Comput.* **2013**, *51*, 593–605.
- [28] Rissanen, S.M.; Kankaanpää, M.; Meigal, A.; Tarvainen, M.P.; Nuutinen, J.; Tarkka, I.M.; Airaksinen, O.; Karjalainen, P.A. Surface EMG and acceleration signals in Parkinson's disease: feature extraction and cluster analysis. *Med. Biol. Eng. Comput.* **2008**, *46*, 849–858.
- [29] Jansen, B.H.; Miller, V.H.; Mavrofrides, D.C.; Stegink Jansen, C.W. Multidimensional EMG-based assessment of walking dynamics. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2003**, *11*, 294–300.
- [30] Rosati, S.; Agostini, V.; Knaflitz, M.; Balestra, G. Muscle activation patterns during gait: A hierarchical clustering analysis. *Biomed. Signal Process. Control* **2017**, *31*, 463–469.
- [31] Agostini, V.; Rosati, S.; Castagneri, C.; Balestra, G.; Knaflitz, M. Clustering analysis of EMG cyclic patterns: A validation study across multiple locomotion pathologies. In Proceedings of the 2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC); IEEE,

- 2017; pp. 1–5.
- [32] Di Nardo, F.; Mengarelli, A.; Maranesi, E.; Burattini, L.; Fioretti, S. Gender differences in the myoelectric activity of lower limb muscles in young healthy subjects during walking. *Biomed. Signal Process. Control* **2015**, *19*, 14–22.
- [33] Di Nardo, F.; Ghetti, G.; Fioretti, S. Assessment of the activation modalities of gastrocnemius lateralis and tibialis anterior during gait: A statistical analysis. *J. Electromyogr. Kinesiol.* **2013**, *23*, 1428–1433.
- [34] Di Nardo, F.; Fioretti, S. Statistical analysis of surface electromyographic signal for the assessment of rectus femoris modalities of activation during gait. *J. Electromyogr. Kinesiol.* **2013**, *23*, 56–61.
- [35] Rosati, S.; Castagneri, C.; Agostini, V.; Knaflitz, M.; Balestra, G. Muscle contractions in cyclic movements: Optimization of CIMAP algorithm. In Proceedings of the Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS; IEEE, 2017; pp. 58–61.
- [36] Kaufman, L.; Rousseeuw, P.J. *Finding groups in data : an introduction to cluster analysis*; Wiley-Interscience, 1990; ISBN 0470317485.
- [37] Demirmen, F. Mathematical Search Procedures in Facies Modeling in Sedimentary Rocks. In; Springer, Boston, MA, 1972; pp. 81–114.
- [38] Bogey, R. a; Barnes, L. a; Perry, J. Computer algorithms to characterize individual subject EMG profiles during gait. *Arch. Phys. Med. Rehabil.* **1992**, *73*, 835–841.
- [39] Agostini, V.; Castagneri, C.; Rosati, S.; Balestra, G.; Knaflitz, M. Extraction of the principal EMG-activations in 100 children during gait : analysis of rectus femoris and vastus medialis. **2017**, 2017.
- [40] Di Nardo, F.; Mengarelli, A.; Burattini, L.; Maranesi, E.; Agostini, V.; Nascimbeni, A.; Knaflitz, M.; Fioretti, S. Normative EMG patterns of ankle muscle co-contractions in school-age children during gait. *Gait Posture* **2016**, *46*, 161–166.
- [41] Figueiredo, E.M.; Ferreira, G.B.; Maia Moreira, R.C.; Kirkwood, R.N.; Fetters, L. Efficacy of ankle-foot orthoses on gait of children with cerebral palsy: systematic review of literature. *Pediatr. Phys. Ther.* **2008**, *20*, 207–23.
- [42] Skaaret, I.; Steen, H.; Huse, A.B.; Holm, I. Comparison of gait with and without ankle-foot orthoses after lower limb surgery in children with unilateral cerebral palsy. *J. Child. Orthop.* **2019**, *13*, 180–189.
- [43] Brehm, M.A.; Harlaar, J.; Schwartz, M. Effect of ankle-foot orthoses on walking efficiency and gait in children with cerebral palsy. *J. Rehabil. Med.* **2008**, *40*, 529–534.
- [44] Willerslev-Olsen, M.; Andersen, J.B.; Sinkjaer, T.; Nielsen, J.B. Sensory feedback to ankle plantar fl[1] M. Willerslev-Olsen, J.B. Andersen, T. Sinkjaer, J.B. Nielsen, Sensory feedback to ankle plantar flexors is not exaggerated during gait in spastic hemiplegic children with cerebral palsy., *J. Neurophysiol.* 111 (2014). *J. Neurophysiol.* **2014**, *111*, 746–54.
- [45] Di Nardo, F.; Strazza, A.; Mengarelli, A.; Ercolani, S.; Morgoni, N.; Burattini, L.; Agostini, V.; Knaflitz, M.; Fioretti, S. Surface EMG patterns for quantification of thigh muscle co-contraction in school-age children: Normative data during walking. *Gait Posture* **2018**, *61*, 25–33.
- [46] Hubley-Kozey, C.L.; Hill, N.A.; Rutherford, D.J.; Dunbar, M.J.; Stanish, W.D. Co-activation differences in lower limb muscles between

- asymptomatic controls and those with varying degrees of knee osteoarthritis during walking. *Clin. Biomech.* **2009**, *24*, 407–414.
- [47] Benedetti, M.G.; Catani, F.; Bilotta, T.W.; Marcacci, M.; Mariani, E.; Giannini, S. Muscle activation pattern and gait biomechanics after total knee replacement. *Clin. Biomech.* **2003**, *18*, 871–876.
- [48] Vaughan, C.L.; Berman, B.; Peacock, W.J. Cerebral palsy and rhizotomy. A 3-year follow-up evaluation with gait analysis. *J. Neurosurg.* **1991**, *74*, 178–184.
- [49] Hakim, S.; Adams, R.D. The special clinical problem of symptomatic hydrocephalus with normal cerebrospinal fluid pressure. Observations on cerebrospinal fluid hydrodynamics. *J. Neurol. Sci.* **1965**, *2*, 307–327.
- [50] Fisher, C.M. Hydrocephalus as a cause of disturbances of gait in the elderly. *Neurology* **1982**, *32*, 1358–1358.
- [51] Graff-Radford, N.R.; Godersky, J.C. Normal-Pressure Hydrocephalus: Onset of Gait Abnormality Before Dementia Predicts Good Surgical Outcome. *Arch. Neurol.* **1986**, *43*, 940–942.
- [52] Malm, J.; Kristensen, B.; Karlsson, T.; Fagerlund, M.; Elfverson, J.; Ekstedt, J. The Predictive Value of Cerebrospinal Fluid Dynamic Tests in Patients With the Idiopathic Adult Hydrocephalus Syndrome. *Arch. Neurol.* **1995**, *52*, 783–789.
- [53] Castagneri, C.; Agostini, V.; Rosati, S.; Balestra, G.; Knaflitz, M. Longitudinal assessment of muscle function after Total Hip Arthroplasty : Use of clustering to extract principal activations from EMG signals. *2018 IEEE Int. Symp. Med. Meas. Appl.* **2018**, *3528725544*, 1–5.
- [54] Higgins, B.T.; Barlow, D.R.; Heagerty, N.E.; Lin, T.J. Anterior vs. Posterior Approach for Total Hip Arthroplasty, a Systematic Review and Meta-analysis. *J. Arthroplasty* **2015**, *30*, 419–434.
- [55] Foucher, K.C. Identifying clinically meaningful benchmarks for gait improvement after total hip arthroplasty. *J. Orthop. Res.* **2016**, *34*, 88–96.
- [56] Bonato, P.; D’Alessio, T.; Knaflitz, M. A statistical method for the measurement of muscle activation intervals from surface myoelectric signal during gait. *IEEE Trans. Biomed. Eng.* **1998**, *45*, 287–299.
- [57] Patterson, K.K.; Gage, W.H.; Brooks, D.; Black, S.E.; McIlroy, W.E. Evaluation of gait symmetry after stroke: A comparison of current methods and recommendations for standardization. *Gait Posture* **2010**, *31*, 241–246.
- [58] Patterson, K.K.; Parafianowicz, I.; Danells, C.J.; Closson, V.; Verrier, M.C.; Staines, W.R.; Black, S.E.; McIlroy, W.E. Gait Asymmetry in Community-Ambulating Stroke Survivors. *Arch. Phys. Med. Rehabil.* **2008**, *89*, 304–310.
- [59] Jorgensen, L.; Crabtree, N.J.; Reeve, J.; Jacobsen, B.K. Ambulatory level and asymmetrical weight bearing after stroke affects bone loss in the upper and lower part of the femoral neck differently: Bone adaptation after decreased mechanical loading. *Bone* **2000**, *27*, 701–707.
- [60] Viteckova, S.; Kutilek, P.; Svoboda, Z.; Krupicka, R.; Kauler, J.; Szabo, Z. Gait symmetry measures: A review of current and prospective methods. *Biomed. Signal Process. Control* **2018**, *42*, 89–100.
- [61] Cabral, S.; Fernandes, R.; Selbie, W.S.; Moniz-Pereira, V.; Veloso, A.P. Inter-session agreement and reliability of the Global Gait Asymmetry index in healthy adults. *Gait Posture* **2017**, *51*, 20–24.
- [62] Auvinet, E.; Multon, F.; Manning, V.; Meunier, J.; Cobb, J.P. Validity and sensitivity of the longitudinal asymmetry index to detect gait asymmetry

- using Microsoft Kinect data. *Gait Posture* **2017**, *51*, 162–168.
- [63] Heredia-Jimenez, J.; Orantes-Gonzalez, E.; Soto-Hermoso, V.M. Variability of gait, bilateral coordination, and asymmetry in women with fibromyalgia. *Gait Posture* **2016**, *45*, 41–44.
- [64] Plate, A.; Sedunko, D.; Pelykh, O.; Schlick, C.; Ilmberger, J.R.; Bötzel, K. Normative data for arm swing asymmetry: How (a)symmetrical are we? *Gait Posture* **2015**, *41*, 13–18.
- [65] Patterson, K.K.; Mansfield, A.; Biasin, L.; Brunton, K.; Inness, E.L.; McIlroy, W.E. Longitudinal changes in poststroke spatiotemporal gait asymmetry over inpatient rehabilitation. *Neurorehabil. Neural Repair* **2015**, *29*, 153–62.
- [66] Kutilek, P.; Viteckova, S.; Svoboda, Z.; Smrcka, P. Kinematic quantification of gait asymmetry in patients with peroneal nerve palsy based on bilateral cyclograms. *J. Musculoskelet. Neuronal Interact.* **2013**, *13*, 244–50.
- [67] Lathrop-Lambach, R.L.; Asay, J.L.; Jamison, S.T.; Pan, X.; Schmitt, L.C.; Blazek, K.; Siston, R.A.; Andriacchi, T.P.; Chaudhari, A.M.W. Evidence for joint moment asymmetry in healthy populations during gait. *Gait Posture* **2014**, *40*, 526–531.
- [68] Schmidt, A.; Stief, F.; Lenarz, K.; Froemel, D.; Lutz, F.; Barker, J.; Meurer, A. Unilateral hip osteoarthritis: Its effects on preoperative lower limb muscle activation and intramuscular coordination patterns. *Gait Posture* **2016**, *45*, 187–192.
- [69] Burnett, D.R.; Campbell-Kyureghyan, N.H.; Cerrito, P.B.; Quesada, P.M. Symmetry of ground reaction forces and muscle activity in asymptomatic subjects during walking, sit-to-stand, and stand-to-sit tasks. *J. Electromyogr. Kinesiol.* **2011**, *21*, 610–615.
- [70] Castagneri, C.; Agostini, V.; Rosati, S.; Balestra, G.; Knaflitz, M. Asymmetry Index in Muscle Activations. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 772–779.
- [71] Veth, R.P.H.; van Hoesel, R.; Pruszczynski, M.; Hoogenhout, J.; Schreuder, B.; Wobbes, T. Limb salvage in musculoskeletal oncology. *Lancet Oncol.* **2003**, *4*, 343–350.
- [72] Höll, S.; Schlomberg, A.; Gosheger, G.; Dieckmann, R.; Streitbuerger, A.; Schulz, D.; Harges, J. Distal femur and proximal tibia replacement with megaprosthesis in revision knee arthroplasty: a limb-saving procedure. *Knee Surgery, Sport. Traumatol. Arthrosc.* **2012**, *20*, 2513–2518.
- [73] Kawai, A.; Backus, S.I.; Otis, J.C.; Inoue, H.; Healey, J.H. Gait characteristics of patients after proximal femoral replacement for malignant bone tumour. *J. Bone Joint Surg. Br.* **2000**, *82-B*, 666–669.
- [74] Øtensjø, S.; Carlberg, E.B.; Vøllestad, N.K. Motor impairments in young children with cerebral palsy: Relationship to gross motor function and everyday activities. *Dev. Med. Child Neurol.* **2004**, *46*, 580–589.
- [75] Graff-Radford, N.R. Normal Pressure Hydrocephalus. *Neurol. Clin.* **2007**, *25*, 809–832.
- [76] Mazzetta, I.; Zampogna, A.; Suppa, A.; Gumiero, A.; Pessione, M.; Irrera, F. Wearable Sensors System for an Improved Analysis of Freezing of Gait in Parkinson's Disease Using Electromyography and Inertial Signals. *Sensors* **2019**, *19*, 948.
- [77] Martínez, M.; Villagra, F.; Castellote, J.M.; Pastor, M.A. Kinematic and kinetic patterns related to free-walking in parkinson's disease. *Sensors*

- (Switzerland) **2018**, 18.
- [78] Khoury, N.; Attal, F.; Amirat, Y.; Oukhellou, L.; Mohammed, S. Data-driven based approach to aid Parkinson's disease diagnosis. *Sensors (Switzerland)* **2019**, 19, 1–27.
- [79] Kamruzzaman, J.; Begg, R.K. Support Vector Machines and Other Pattern Recognition Approaches to the Diagnosis of Cerebral Palsy Gait. *IEEE Trans. Biomed. Eng.* **2006**, 53, 2479–2490.
- [80] Zhang, B.; Zhang, Y.; Begg, R.K. Gait classification in children with cerebral palsy by Bayesian approach. *Pattern Recognit.* **2009**, 42, 581–586.
- [81] Kugler, P.; Jaremenko, C.; Schlachetzki, J.; Winkler, J.; Klucken, J.; Eskofier, B. Automatic recognition of Parkinson's disease using surface electromyography during standardized gait tests. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); IEEE, 2013; pp. 5781–5784.
- [82] Bojanic, D.M.; Petrovacki-Balj, B.D.; Jorgovanovic, N.D.; Ilic, V.R. Quantification of dynamic EMG patterns during gait in children with cerebral palsy. *J. Neurosci. Methods* **2011**, 198, 325–331.
- [83] Vargas, J.M. The Probabilistic Basis of Jaccard's Index of Similarity. *Artic. Syst. Biol.* **1996**.
- [84] Schwartz, M.H.; Rozumalski, A.; Trost, J.P. The effect of walking speed on the gait of typically developing children. **2008**, 41, 1639–1650.
- [85] Sutherland, D.H.; Olshen, R.; Cooper, L.; Woo, S.L. The development of mature gait. *J. Bone Joint Surg. Am.* **1980**, 62, 336–53.
- [86] Chang, W.N.; Lipton, J.S.; Tsirikos, A.I.; Miller, F. Kinesiological surface electromyography in normal children: Range of normal activity and pattern analysis. *J. Electromyogr. Kinesiol.* **2007**, 17, 437–445.
- [87] Cheung, V.C.K.; d'Avella, A.; Tresch, M.C.; Bizzi, E. Central and sensory contributions to the activation and organization of muscle synergies during natural motor behaviors. *J. Neurosci.* **2005**, 25, 6419–34.
- [88] Chvatal, S.A.; Ting, L.H. Common muscle synergies for balance and walking. *Front. Comput. Neurosci.* **2013**.
- [89] Hagio, S.; Fukuda, M.; Kouzaki, M. Identification of muscle synergies associated with gait transition in humans. *Front. Hum. Neurosci.* **2015**, 9.
- [90] Ivanenko, Y.P.; Poppele, R.E.; Lacquaniti, F. Five basic muscle activation patterns account for muscle activity during human locomotion. *J. Physiol.* **2004**, 556, 267–282.
- [91] Lee, D.D.; Seung, H.S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, 401, 788–791.
- [92] Torres-Oviedo, G.; Ting, L.H. Muscle synergies characterizing human postural responses. *J. Neurophysiol.* **2007**, 98, 2144–56.
- [93] Rimini, D.; Agostini, V.; Rosati, S.; Castagneri, C.; Balestra, G.; Knaflitz, M. Influence of pre-processing in the extraction of muscle synergies during human locomotion. In Proceedings of the Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS; Institute of Electrical and Electronics Engineers Inc., 2017; pp. 2502–2505.
- [94] Ghislieri, M.; Agostini, V.; Knaflitz, M. How to Improve Robustness in Muscle Synergy Extraction.; Institute of Electrical and Electronics Engineers (IEEE), 2019; pp. 1525–1528.
- [95] Zelik, K.E.; La Scaleia, V.; Ivanenko, Y.P.; Lacquaniti, F. Can modular strategies simplify neural control of multidirectional human locomotion? *J.*

- Neurophysiol.* **2014**, *111*, 1686–702.
- [96] Ghislieri, M.; Agostini, V.; Knaflitz, M. Muscle Synergies Extracted Using Principal Activations: Improvement of Robustness and Interpretability. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 453–460.
- [97] D’Avella, A.; Bizzi, E. Shared and specific muscle synergies in natural motor behaviors. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 3076–3081.
- [98] Bizzi, E.; Cheung, V.C.K. The neural origin of muscle synergies. *Front. Comput. Neurosci.* **2013**, *7*, 1–6.
- [99] Rimini, D.; Agostini, V.; Knaflitz, M. Intra-subject consistency during locomotion: Similarity in shared and subject-specific muscle synergies. *Front. Hum. Neurosci.* **2017**, *11*.
- [100] Clark, D.J.; Ting, L.H.; Zajac, F.E.; Neptune, R.R.; Kautz, S.A. Merging of healthy motor modules predicts reduced locomotor performance and muscle coordination complexity post-stroke. *J. Neurophysiol.* **2010**, *103*, 844–857.
- [101] Di Nardo, F.; Mengarelli, A.; Strazza, A.; Agostini, V.; Knaflitz, M.; Burattini, L.; Fioretti, S. A new parameter for quantifying the variability of surface electromyographic signals during gait: The occurrence frequency. *J. Electromyogr. Kinesiol.* **2017**, *36*, 25–33.
- [102] Rodriguez, K.L.; Roemmich, R.T.; Cam, B.; Fregly, B.J.; Hass, C.J. Persons with Parkinson’s disease exhibit decreased neuromuscular complexity during gait. *Clin. Neurophysiol.* **2013**, *124*, 1390–7.
- [103] Allen, J.L.; McKay, J.L.; Sawers, A.; Hackney, M.E.; Ting, L.H. Increased neuromuscular consistency in gait and balance after partnered, dance-based rehabilitation in Parkinson’s disease. *J. Neurophysiol.* **2017**, *118*, 363–373.

Appendix A: Dataset description

In this section, a complete description of the datasets used in the thesis is provided. All the data were previously collected for several studies and, in this thesis, data were extracted from existing datasets and retrospectively analyzed. The datasets can be divided into two macro-categories: *healthy subjects* and *pathological subjects*. Dataset details are reported in Table A.1.

Healthy subjects

- **School-age children**

This dataset consists of gait data of 100 healthy school-age children (51 males / 49 females) [14] collected at ASL TO4 (Moncalieri, TO).

The entire dataset was used for the study *Characterization of school-age children population* (Chapter 4.3). Moreover, a subset of 10 children was extracted for the study related to the *EMG Asymmetry Index* (Chapter 5.2). Finally, a subset consisting of 80 children was extracted and used as Validation Set in the *Muscle Functionality Index* study (Chapter 5.3.2).

- **Adult and elderly controls**

In previous studies, three main datasets were collected consisting of gait data of healthy adult and elderly.

The first dataset refers to the study presented in [15] and consists of 18 healthy subjects (8 males / 10 females). Gait data were collected at BIOLAB laboratory (Politecnico di Torino).

The entire dataset was used as part of the *Test Set* for testing the CIMAP algorithm (Chapter 2.3). Moreover, a subset of 6 subjects was extracted for the study related to the *EMG Asymmetry Index* and used in the Adult control group (Chapter 5.2).

The second dataset refers to the study presented in [53] and consists of 20 healthy subjects (11 males / 9 females). Gait data were collected at Rehabilitation and Functional Recovery Unit at the Ivrea Hospital, Torino (Italy).

The entire dataset was used for the study *Assessment of muscle activation differences between idiopathic normal pressure hydrocephalus patients and healthy controls* (Chapter 4.4) and for the study *Longitudinal assessment of muscle function in patients with Total Hip Arthroplasty* (Chapter 4.5). Moreover, a subset of 5 subjects was extracted and used as part of the *Validation_opt* dataset for validating the

optimized version of the CIMAP algorithm (Chapter 3.5). Finally, a subset of 4 subjects was extracted for the study related to the *EMG Asymmetry Index* and used in the Adult control group (Chapter 5.2).

Finally, the third dataset consists of 80 healthy subjects which were enrolled in several pilot study during the years.

Gait data of 17 subjects were extracted from this dataset and used in this thesis. Two subjects were used as part of the *Test Set* for testing the CIMAP algorithm (Chapter 2.3), 5 subjects was extracted and used as part of the *Validation_opt* dataset for validating the optimized version of the CIMAP algorithm (Chapter 3.5) and 10 subjects were used for the study *Assessment of muscle activation differences between idiopathic normal pressure hydrocephalus patients and healthy controls* (Chapter 4.4)

Pathological subjects

- **Hemiplegic children**

This dataset consists of 25 children (15 males / 10 females) with Winters' type I and II hemiplegia after cerebral palsy. Gait data were collected at ASL TO4 (Moncalieri, TO).

The entire dataset was used for the study related to the *EMG Asymmetry Index* (Chapter 5.2) and for *Muscle Functionality Index* study (Chapter 5.3.2).

- **THA patients**

This dataset refers to the study presented in [53] and consists of 20 patients (9 males / 11 females) that underwent a Total Hip Arthroplasty (THA) procedure. They were recruited from the Rehabilitation and Functional Recovery Unit at the Ivrea Hospital, Torino.

The entire dataset was used for the study *Longitudinal assessment of muscle function in patients with Total Hip Arthroplasty* (Chapter 4.5) and for the study related to the *EMG Asymmetry Index* (Chapter 5.2).

Moreover, a subset of 5 subjects was extracted and used as part of the *Validation Set* for validating the first version of CIMAP algorithm (Chapter 2.4) and 5 subjects were extracted and used as part of the *Validation_opt* dataset for validating the optimized version of the CIMAP algorithm (Chapter 3.5).

- **TKR patients**

This dataset consists of 10 patients (5 males / 5 females) that underwent a Total Knee Replacement (TKR) procedure. Gait data were collected at "Città della Salute e della Scienza" (Molinette Hospital, Torino).

The entire dataset was used for the study related to the *EMG Asymmetry Index* (Chapter 5.2). Moreover, a subset of 5 subjects was extracted and used as part of

the *Validation Set* for validating the first version of CIMAP algorithm (Chapter 2.4) and 5 subjects were extracted and used as part of the *Validation_opt* dataset for validating the optimized version of the CIMAP algorithm (Chapter 3.5).

- **Mega TKR patients**

The subjects included in the Mega TKR dataset are patients affected by malignant tumours of the distal part of the femur, who underwent the surgical procedure of modular knee prosthesis (megaprosthesis) implant for saving the lower limb. This dataset consists of 19 patients (10 males / 9 females). Gait data were collected at “Città della Salute e della Scienza” (Molinette Hospital, Torino).

The entire dataset was used for the study related to the *EMG Asymmetry Index* (Chapter 5.2). Moreover, a subset of 5 subjects was extracted and used as part of the *Validation Set* for validating the first version of CIMAP algorithm (Chapter 2.4) and 5 subjects were extracted and used as part of the *Validation_opt* dataset for validating the optimized version of the CIMAP algorithm (Chapter 3.5).

- **iNPH patients**

This dataset consists of 30 patients (25 males / 5 females) affected by idiopathic Normal Pressure Hydrocephalus (iNPH). Gait data were collected at “Città della Salute e della Scienza” (Molinette Hospital, Torino).

The entire dataset was used for the study *Assessment of muscle activation differences between idiopathic normal pressure hydrocephalus patients and healthy controls* (Chapter 4.4) and a subset of 10 patients was extracted for the study related to the *EMG Asymmetry Index* (Chapter 5.2). Moreover, a subset of 5 subjects was extracted and used as part of the *Validation Set* for validating the first version of CIMAP algorithm (Chapter 2.4) and 5 subjects were extracted and used as part of the *Validation_opt* dataset for validating the optimized version of the CIMAP algorithm (Chapter 3.5).

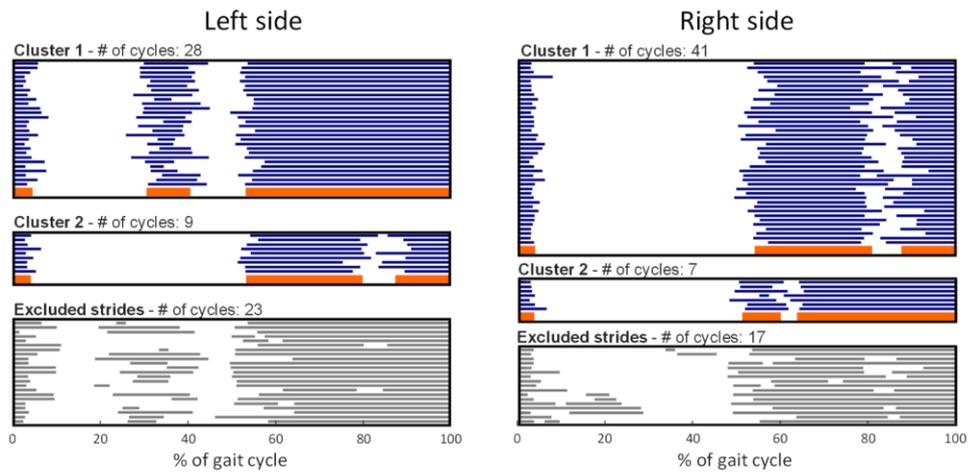
Table A.1. Dataset details. Age, height and weight are reported as mean \pm standard deviation.

Dataset	Anthropometric characteristics		
	Age [years]	Height [cm]	Weight [kg]
<i>School-age children</i>	9.0 \pm 1.4	133.40 \pm 9.3	30.6 \pm 6.7
<i>Adult and elderly controls (First dataset)</i>	24.0 \pm 2.7	169.5 \pm 8.4	63.3 \pm 10
<i>Adult and elderly controls (Second dataset)</i>	65.4 \pm 5.1	169.8 \pm 9.4	69.0 \pm 12.2
<i>Adult and elderly controls (Third dataset)</i>	40.9 \pm 18.1	171.1 \pm 8.0	65.9 \pm 10.8
<i>Hemiplegic children</i>	8.7 \pm 3.2	129.7 \pm 18.8	30.2 \pm 11.7
<i>THA patients</i>	66.1 \pm 7.2	168.7 \pm 10.5	77.0 \pm 13.3
<i>TKR patients</i>	71.2 \pm 8.8	168.2 \pm 9.1	86.8 \pm 21.0
<i>Mega TKR patients</i>	37.8 \pm 17.8	170.4 \pm 10.5	68.9 \pm 11.4
<i>iNPH</i>	73.8 \pm 8.6	167.6 \pm 7.6	72.0 \pm 11.0

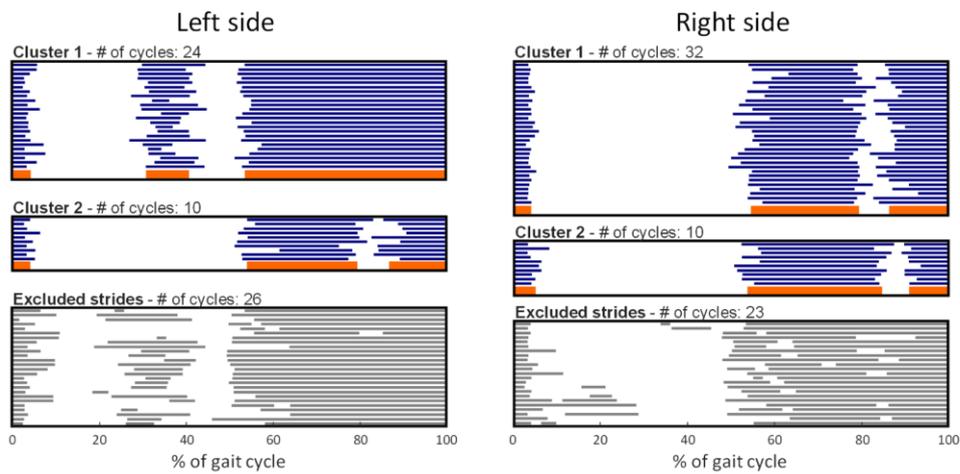
Appendix B: Distance metric tests

In this section, some example of clustering results obtained using three different distance metrics are reported (*euclidean*, *chebyshev* and *cityblock*). Both healthy and pathological subject were used for assessing the effect of the distance metric on the clustering result. The tests were performed on the Tibialis Anterior muscle.

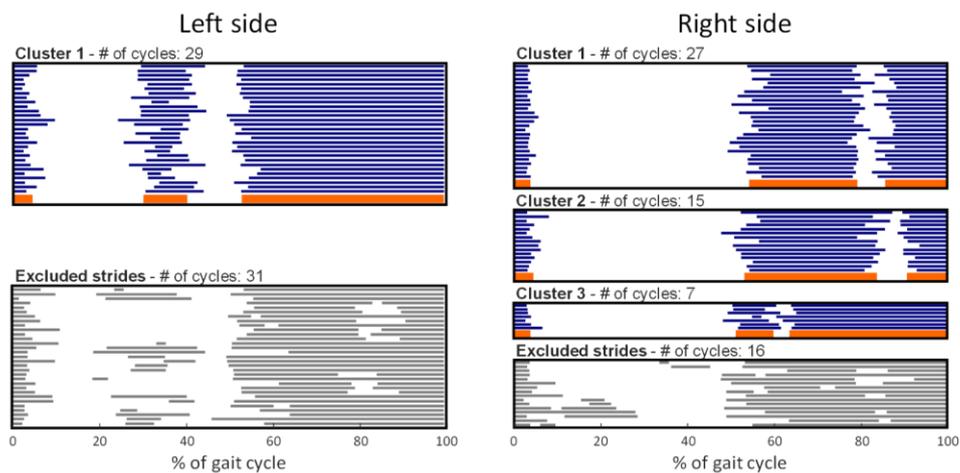
Overall, the *chebyshev* and *cityblock* distances have shown to be the most promising distance metric, while the *euclidean* distance were discarded at the end of the empirical tests.



(a) Chebychev



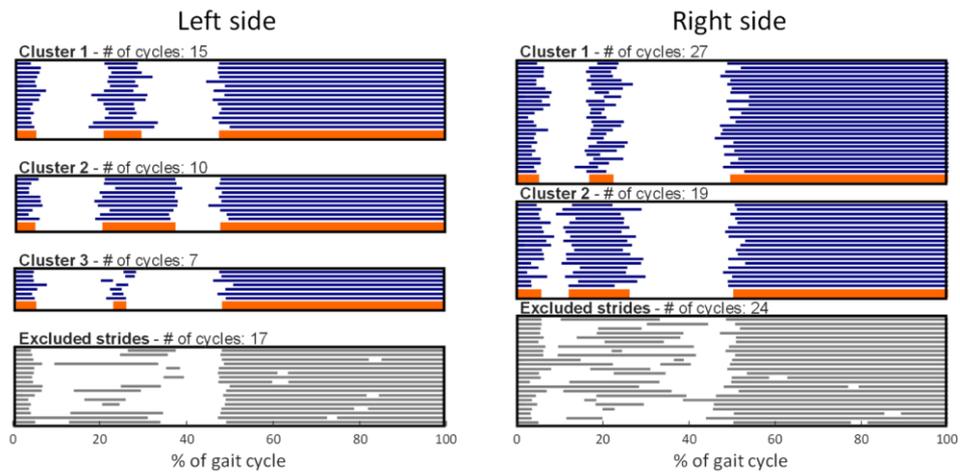
(b) Cityblock



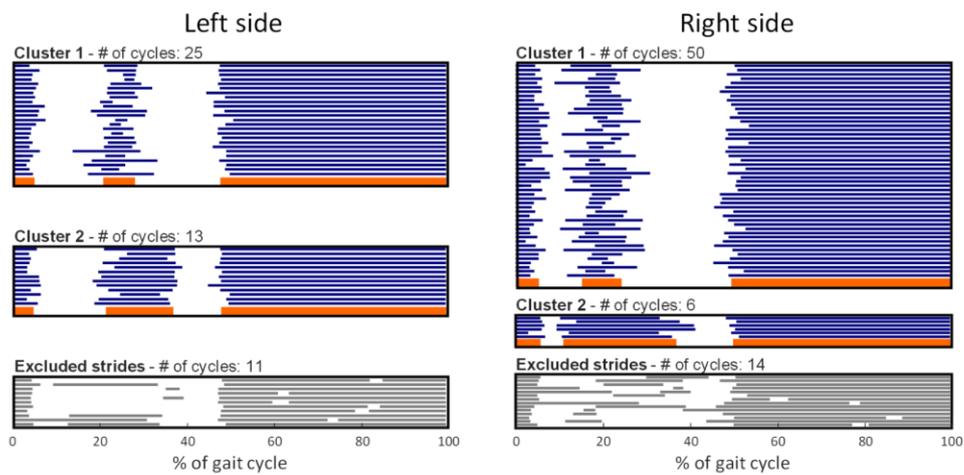
(c) Euclidean

Figure B.1. Clustering results (Healthy male, 25 years old)

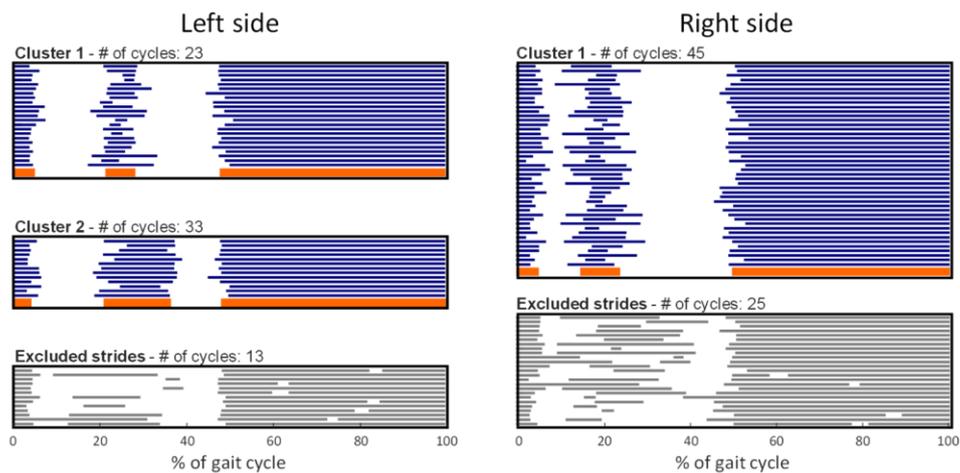
Using Euclidean norm, on the left side only one activation pattern is identified and on the right side two clusters are identified with similar activation patterns. Overall, the best result is obtained using Chebyshev distance: the main activation patterns are identified, with a smaller number of excluded strides with respect to the Cityblock norm. Blue lines represent strides belonging to significant clusters, orange lines represent the centroids of each cluster and grey lines represents strides excluded from significant clusters.



(a) Chebychev



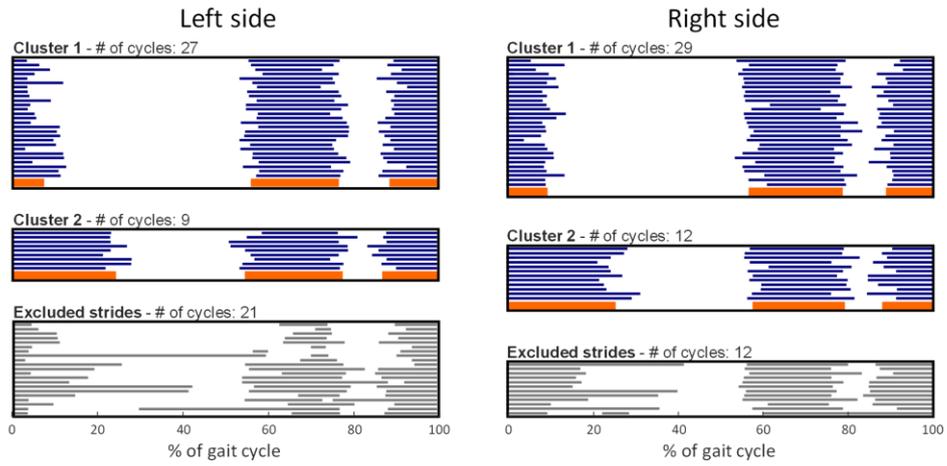
(b) Cityblock



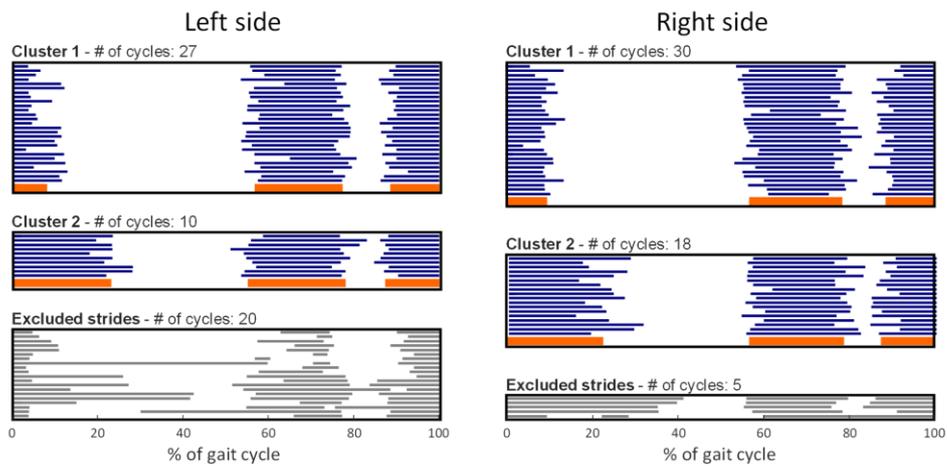
(c) Euclidean

Figure B.2. Clustering results (Healthy male, 30 years old)

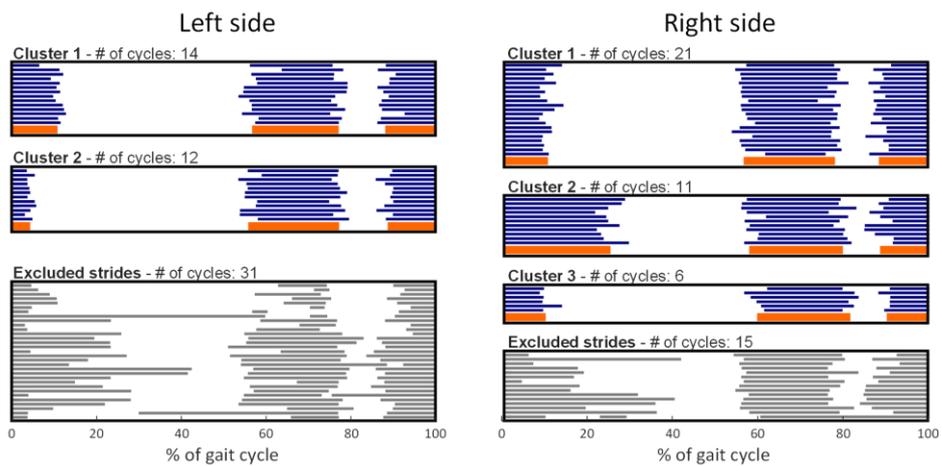
Using Euclidean norm, on the right side only one activation pattern is identified. Overall, the best result is obtained using Cityblock distance: the main activation patterns are identified, with a smaller number of excluded strides with respect to the Chebyshev norm. Blue lines represent strides belonging to significant clusters, orange lines represent the centroids of each cluster and grey lines represent strides excluded from significant clusters.



(a) Chebychev



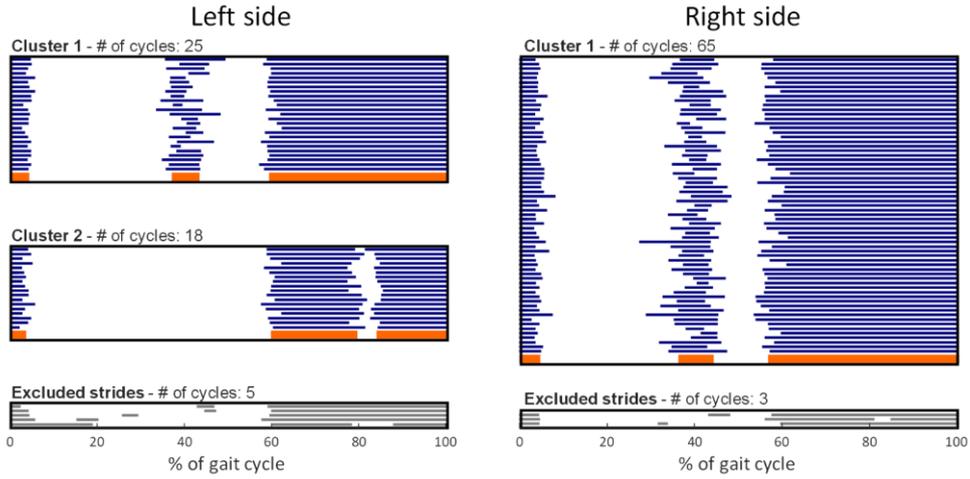
(b) Cityblock



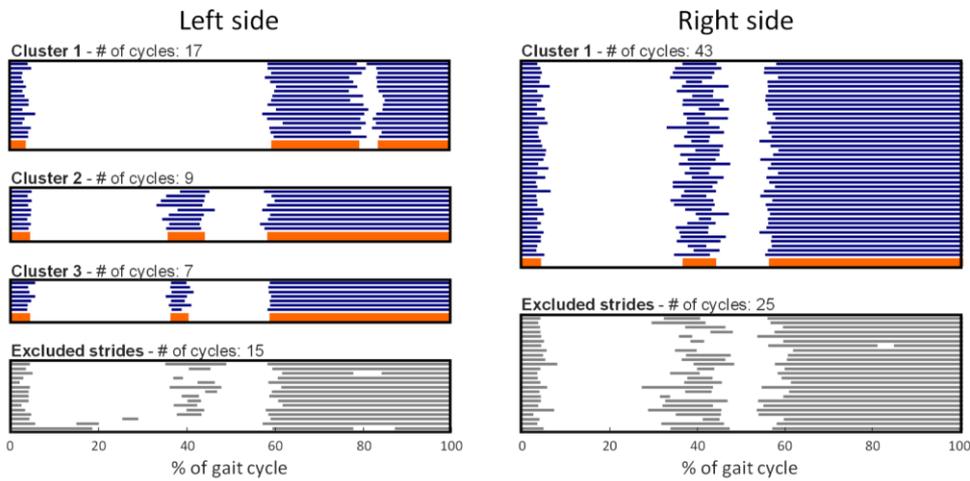
(c) Euclidean

Figure B.3. Clustering results (Healthy female, 28 years old)

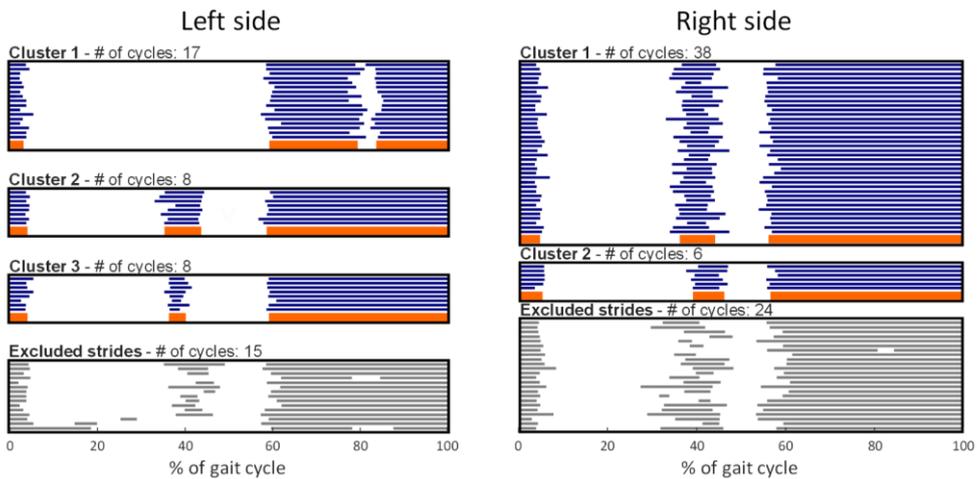
Using Euclidean norm, on the right side, two clusters with similar activation patterns are identified (cluster 1 and cluster 3). Overall, the best result is obtained using Cityblock distance: the main activation patterns are identified, with a smaller number of excluded strides with respect to the Chebychev norm. Blue lines represent strides belonging to significant clusters, orange lines represent the centroids of each cluster and grey lines represents strides excluded from significant clusters.



(a) Chebychev



(b) Cityblock



(c) Euclidean

Figure B.4. Clustering results (Healthy male, 26 years old)

Using Euclidean norm, on both sides, clusters with similar activation patterns are identified (cluster 2 and 3 for the left side, cluster 1 and 2 for the right side). Overall, the best result is obtained using Chebyshev distance: the main activation patterns are identified, with the smaller number of excluded strides with respect to the other norms. Blue lines represent strides belonging to significant clusters, orange lines represent the centroids of each cluster and grey lines represents strides excluded from significant clusters.

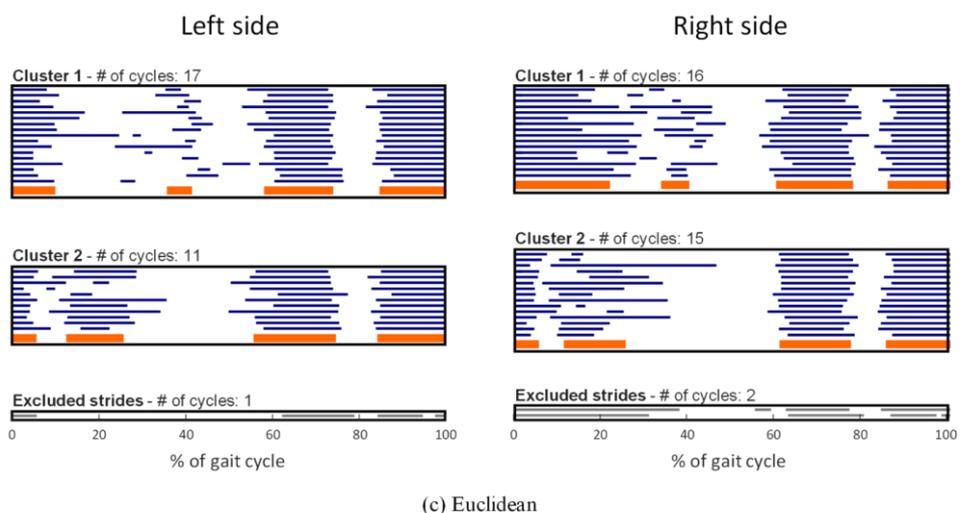
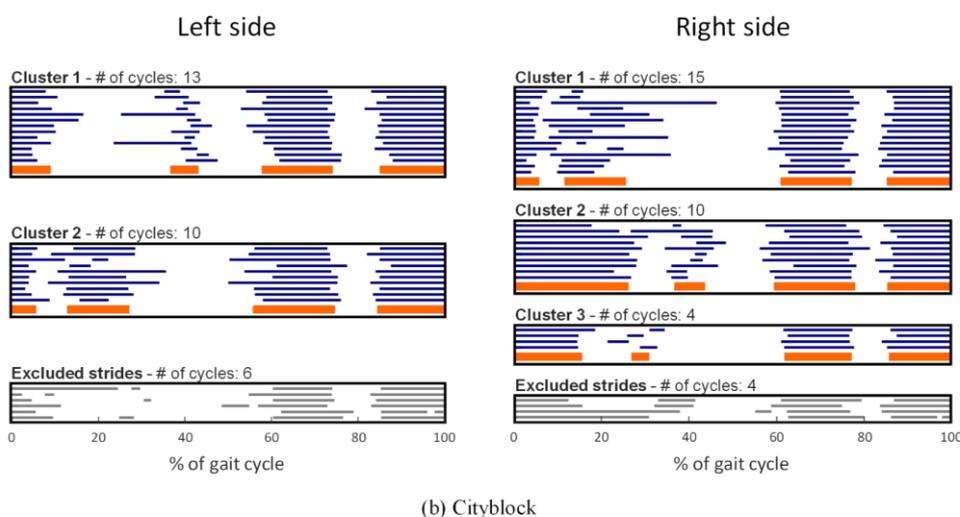
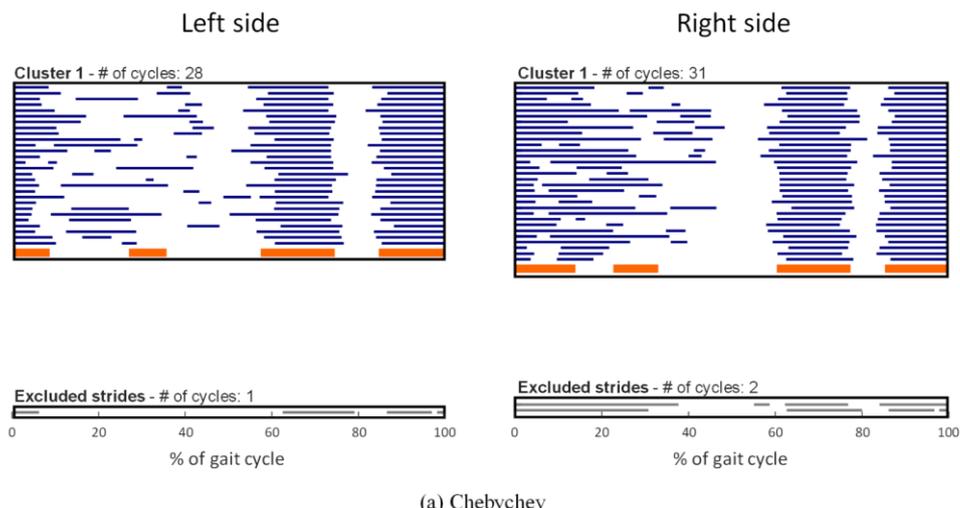
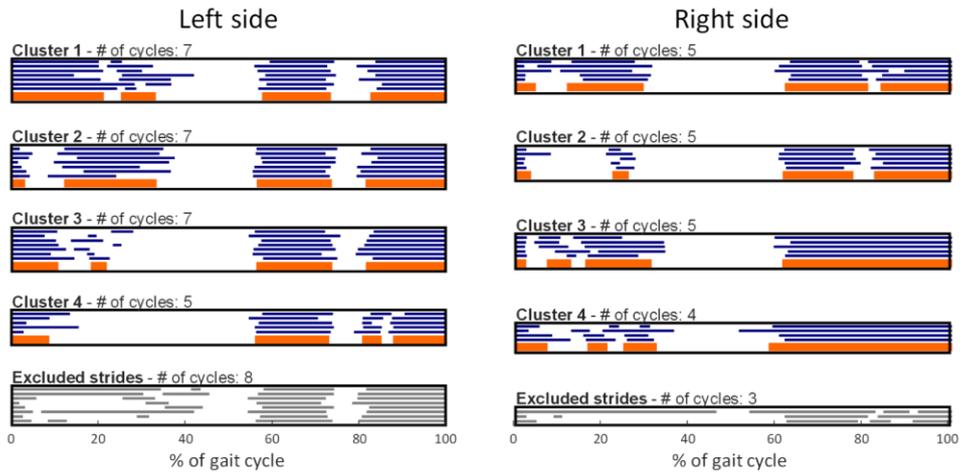
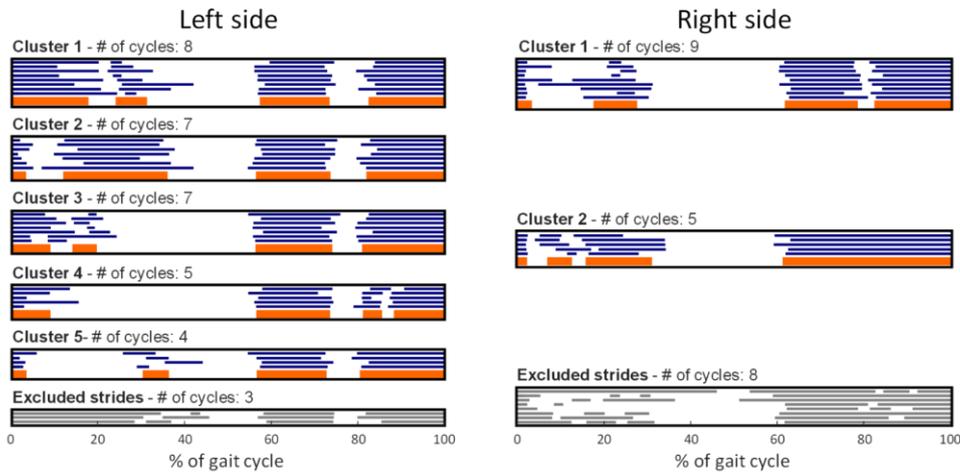


Figure B.5. Clustering results (THA patient, female, 70 years old)

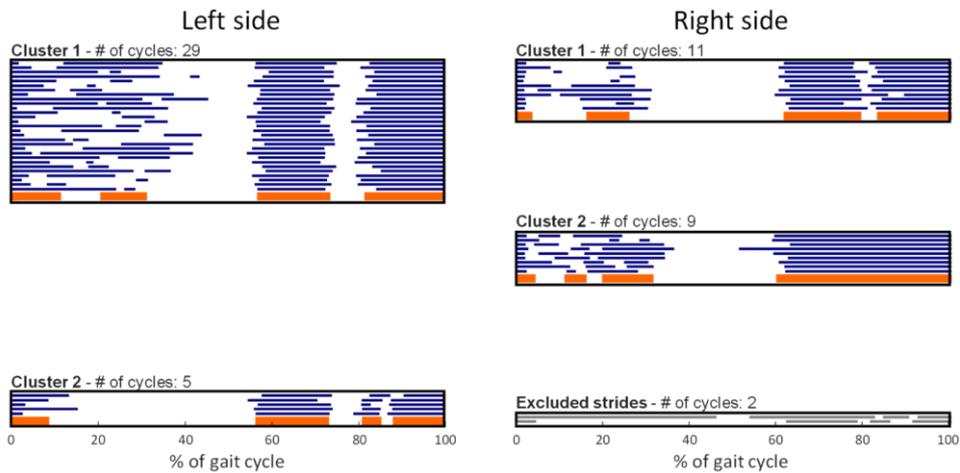
Using Chebyshev norm, both on left and right side the clusterization results not satisfying. The Cityblock and the Euclidean norms instead, allow to identify several different activation patterns. However, the best result is obtained using Cityblock distance: on the right side it allows to identify one more significant cluster with respect to the Euclidean distance. Blue lines represent strides belonging to significant clusters, orange lines represent the centroids of each cluster and grey lines represent strides excluded from significant clusters.



(a) Chebychev



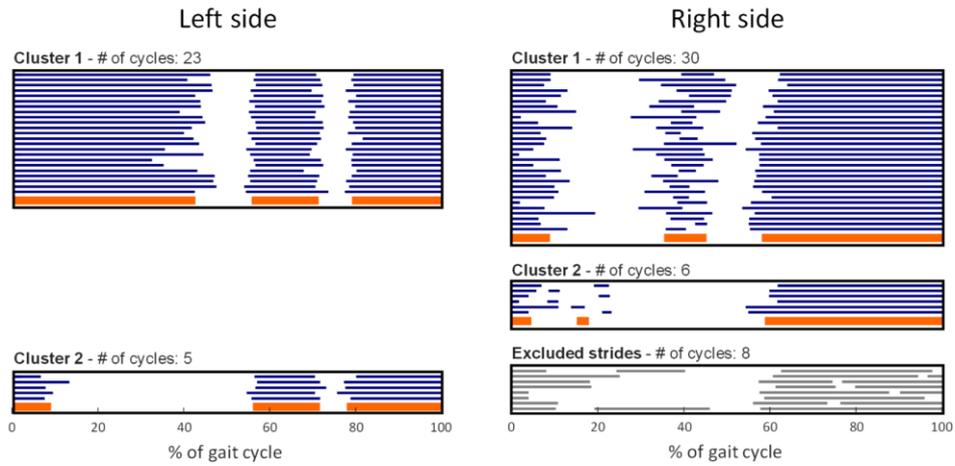
(b) Cityblock



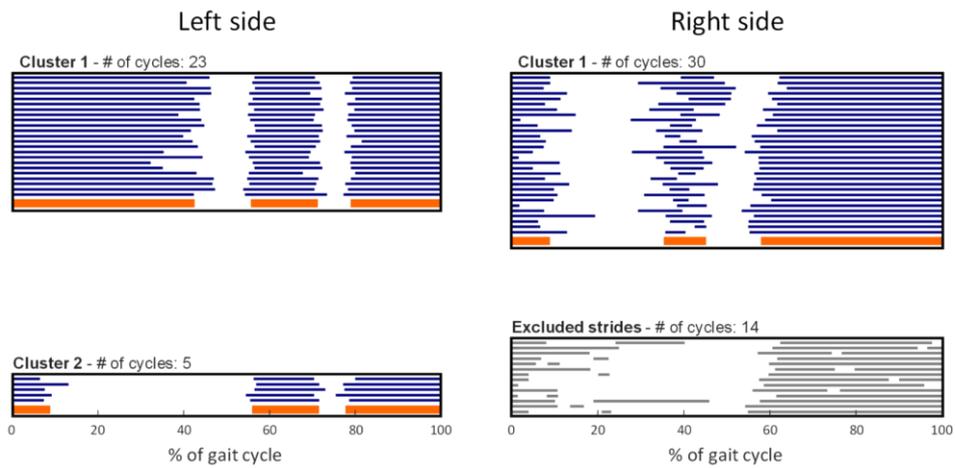
(c) Euclidean

Figure B.6. Clustering results (THA patient, female, 61 years old)

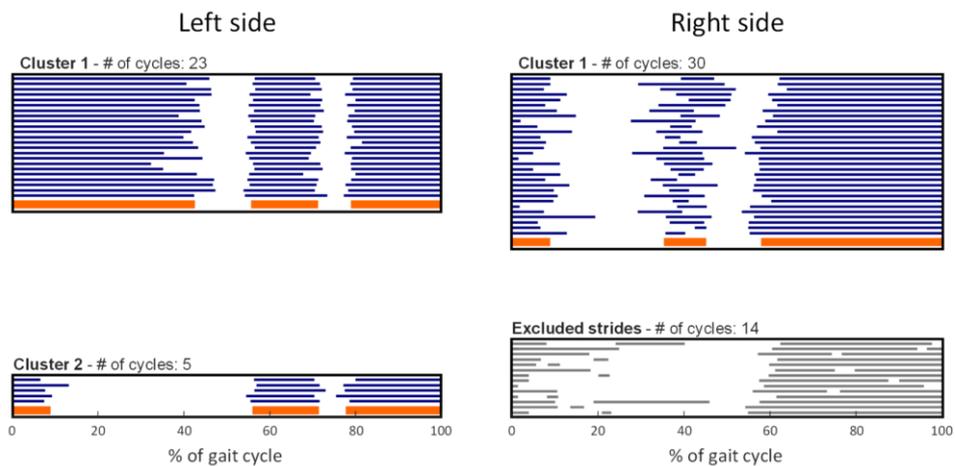
Using Euclidean norm, particularly on the left, the clusterization results not satisfying. The Cityblock and the Chebyshev norms instead, allow to identify several different activation patterns. Overall, the best result is obtained using Chebyshev distance: it allows to identify the main significant clusters on both sides. Blue lines represent strides belonging to significant clusters, orange lines represent the centroids of each cluster and grey lines represents strides excluded from significant clusters.



(a) Chebychev



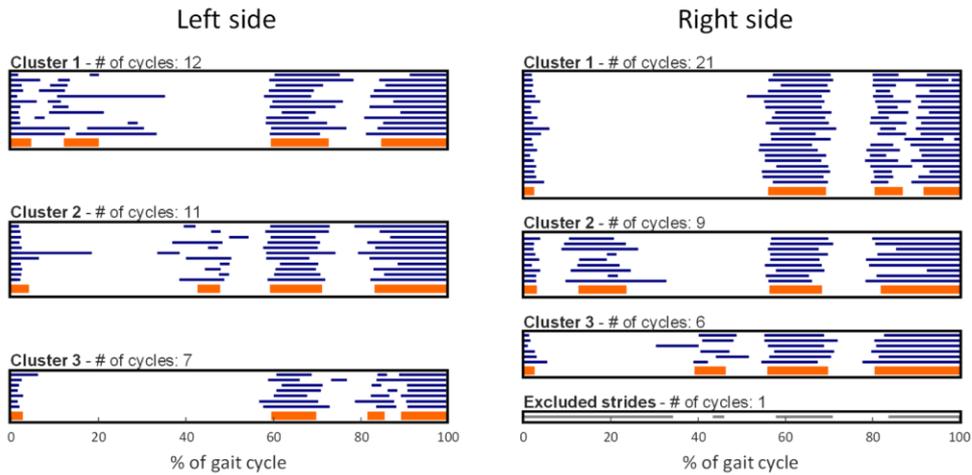
(b) Cityblock



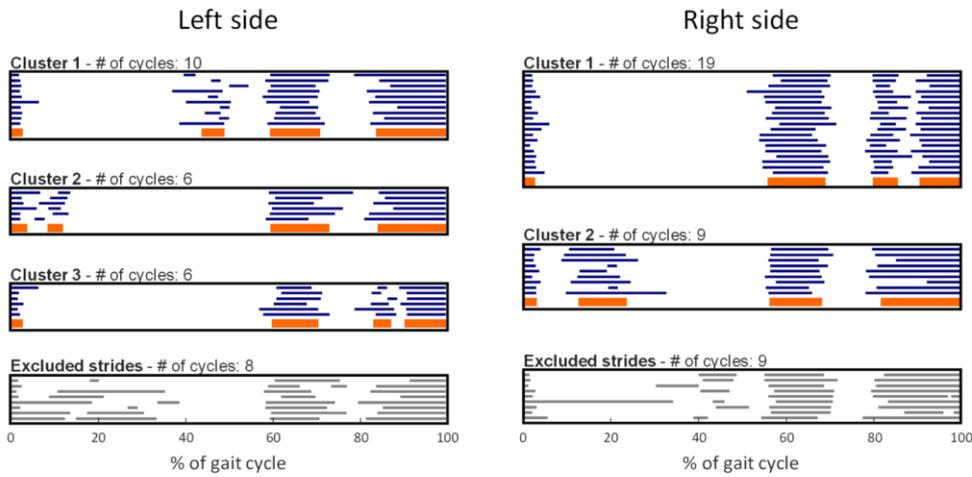
(c) Euclidean

Figure B.7. Clustering results (Mega TKR patient, male, 20 years old)

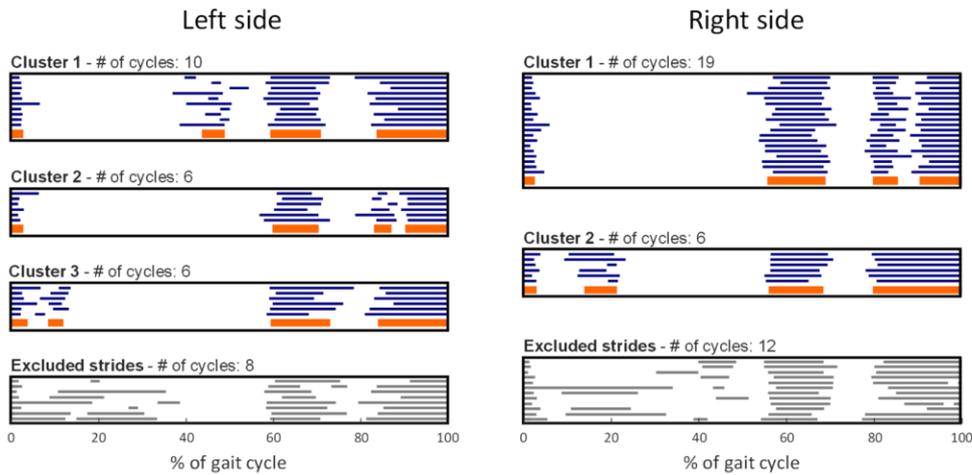
On the left side, the same result is obtained using the three norms. On the right side, the best results are obtained using the Chebyshev distance: it allows to identify the different activation patterns. Blue lines represent strides belonging to significant clusters, orange lines represent the centroids of each cluster and grey lines represents strides excluded from significant clusters.



(a) Chebychev



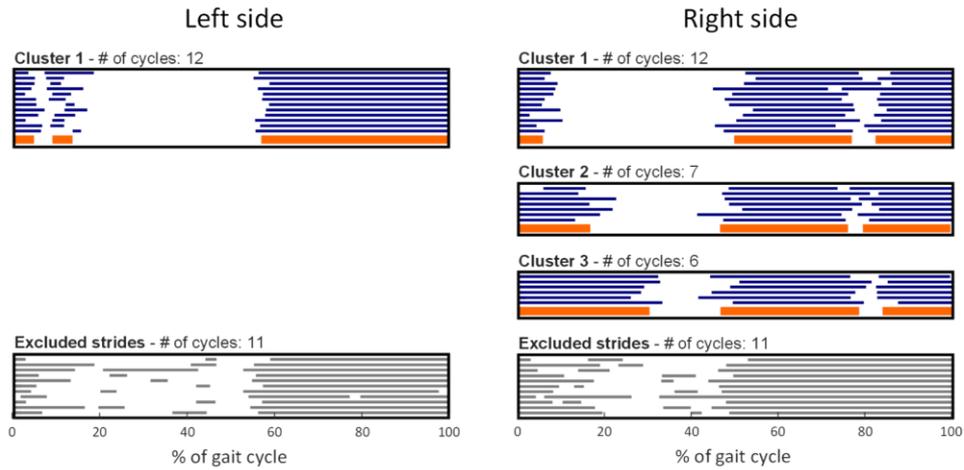
(b) Cityblock



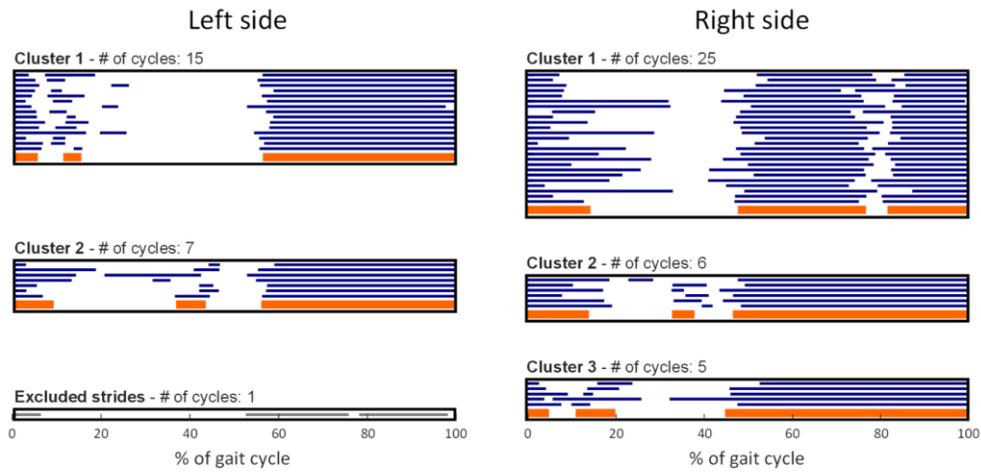
(c) Euclidean

Figure B.8. Clustering results (Mega TKR patient, male, 39 years old)

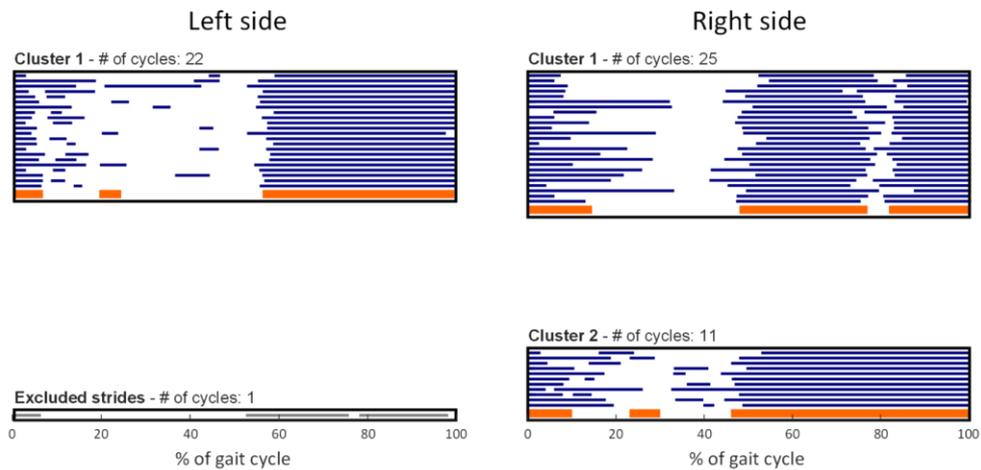
On both sides, almost the same results are obtained using the Cityblock and the Euclidean norms. The best results are obtained using the Chebyshev distance: it allows to identify the main activation patterns on both sides, with the smallest number of excluded strides. Blue lines represent strides belonging to significant clusters, orange lines represent the centroids of each cluster and grey lines represent strides excluded from significant clusters.



(a) Chebychev



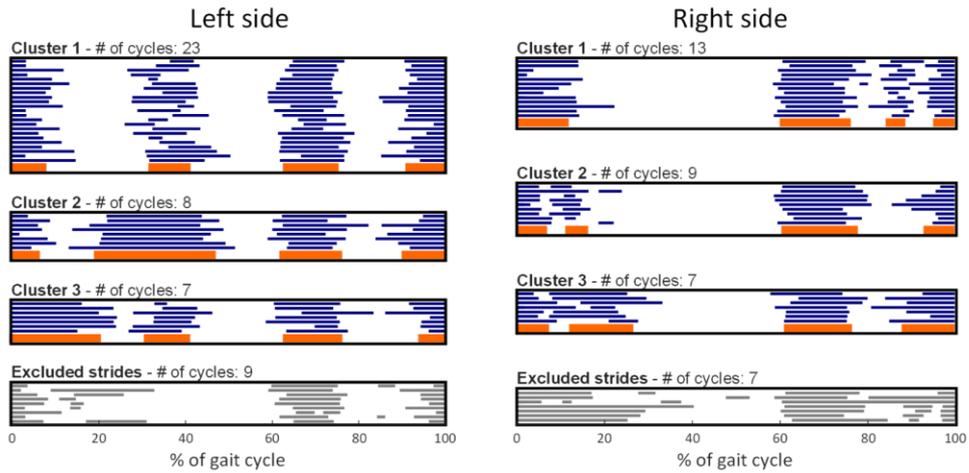
(b) Cityblock



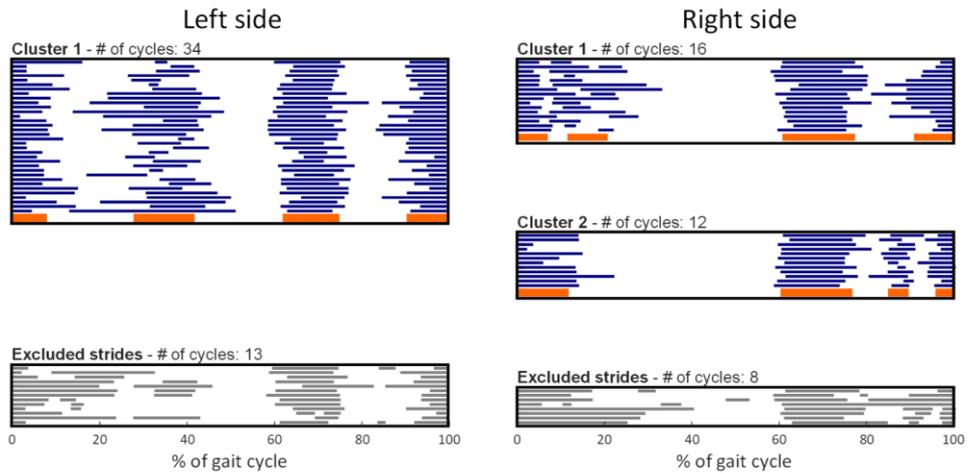
(c) Euclidean

Figure B.9. Clustering results (TKR patient, female, 77 years old)

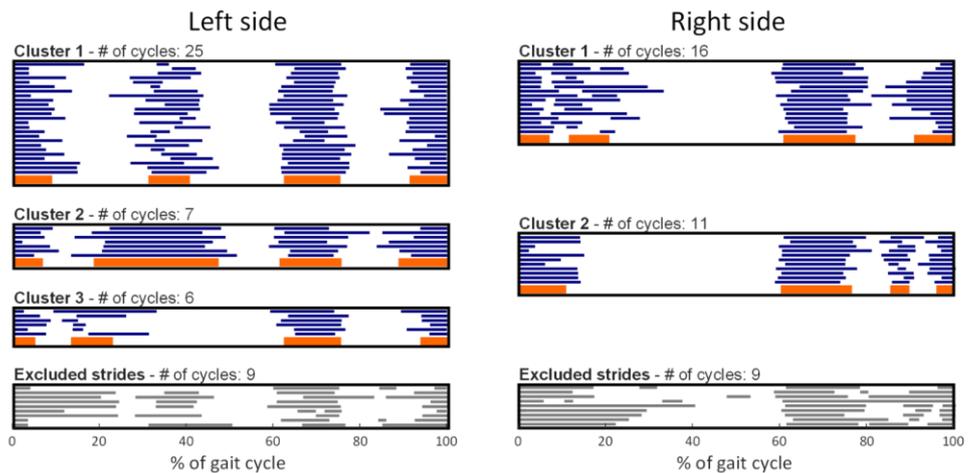
Using the Euclidean and the Chebyshev norms, on both sides, some activation patterns are not identified as significant. The best result is obtained using the Cityblock distance: it allows to identify the main activation patterns on both sides, with only one excluded stride on the left side. Blue lines represent strides belonging to significant clusters, orange lines represent the centroids of each cluster and grey lines represent strides excluded from significant clusters.



(a) Chebychev



(b) Cityblock



(c) Euclidean

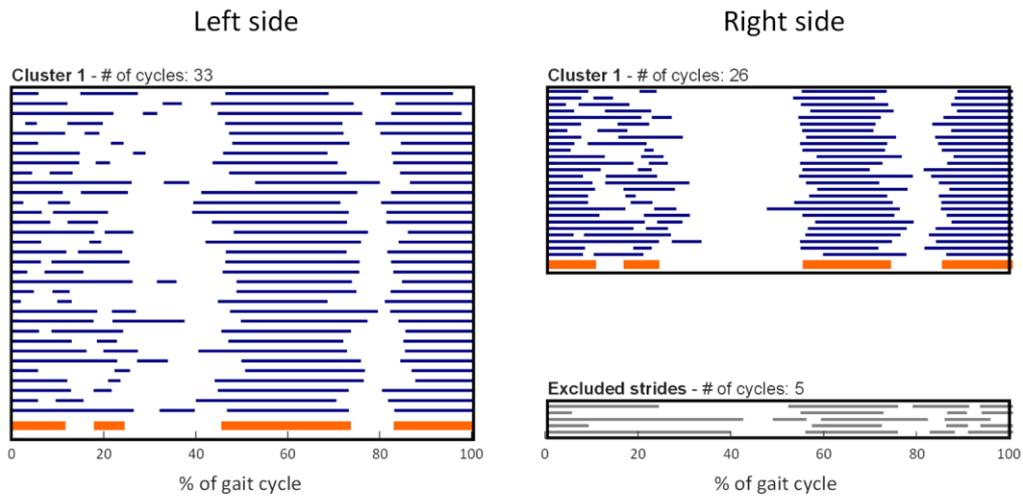
Figure B.10. Clustering results (TKR patient, female, 59 years old)

Using the Euclidean and the Chebyshev norms, similar results are obtained on both sides. However, the best result is obtained using the Chebyshev distance: on the right side, it allows to identify all the main activation patterns. Blue lines represent strides belonging to significant clusters, orange lines represent the centroids of each cluster and grey lines represents strides excluded from significant clusters.

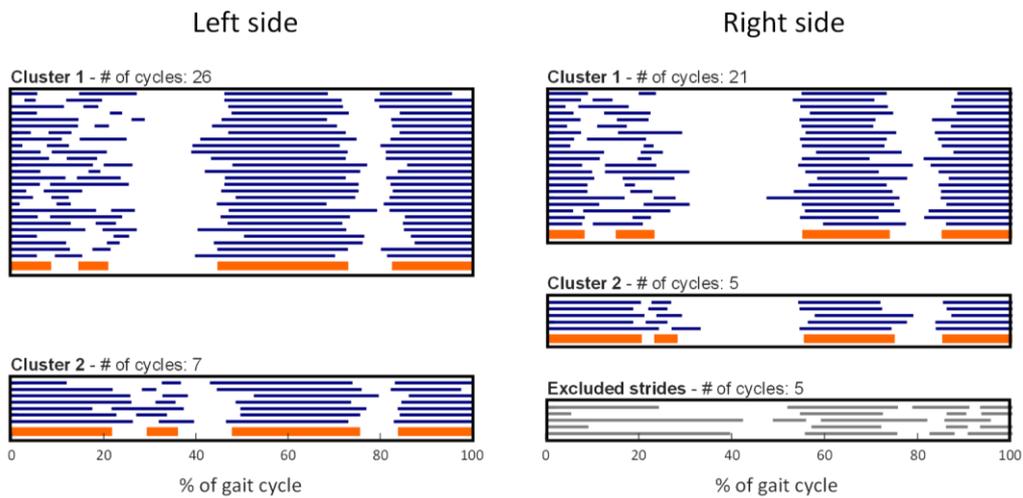
Appendix C: Cutoff point tests

In this section, some example of clustering results obtained cutting the dendrogram tree at iteration k , (see eq. 2.4) are presented. For each example, the result obtained cutting the tree at a different iteration ($k-1$, $k-2$, $k+1$ or $k+2$) are reported. Both healthy and pathological subject were used for assessing the efficacy of the cutting rule based on R_k index. The tests were performed on the Tibialis Anterior muscle.

Overall, it emerges that the cutting rule defined considering the R_k indicator does not allow for obtaining the optimal number of clusters in several situations.



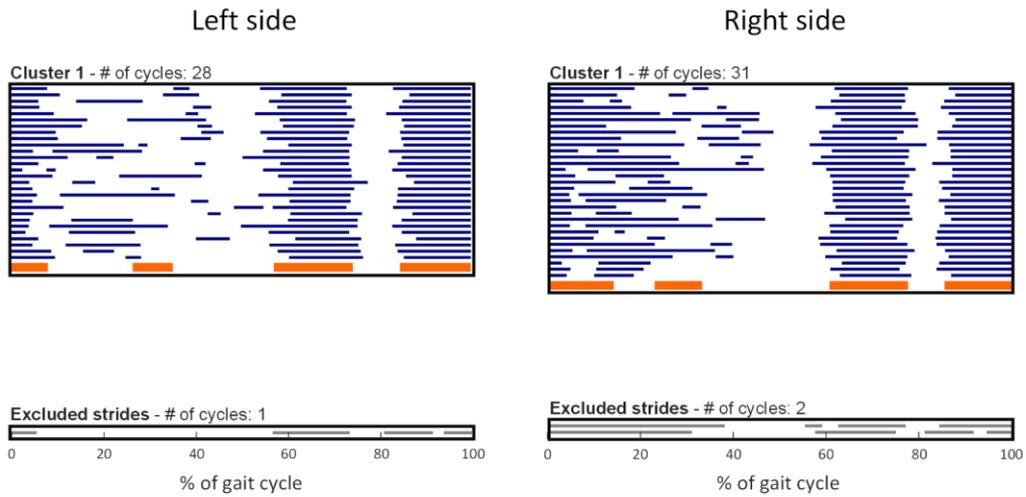
(a) Clustering results – iteration k



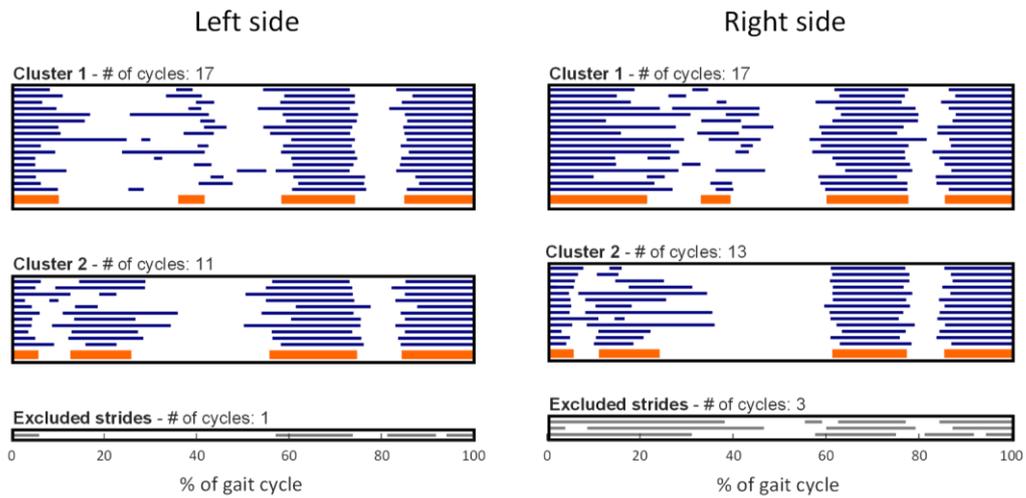
(b) Clustering results – iteration $k-1$

Figure C.1. Clustering results (iNPH patient, female, 85 years old)

Clustering results obtained cutting the dendrogram at (a) iteration k and (b) $k-1$. As it can be observed from panel (a), the clusters obtained cutting the tree at the iteration k are not an optimal result. In panel (b) the clusters obtained cutting the tree at iteration $k-1$ are represented: in this case, both on the left and right side, two different activation patterns can be recognized, while in panel (a) only one cluster per side is identified.



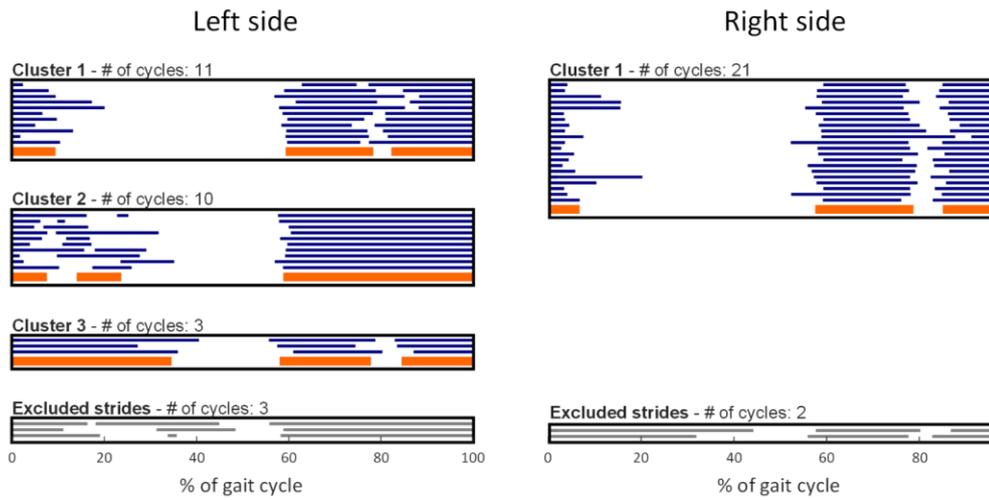
(a) Clustering results – iteration k



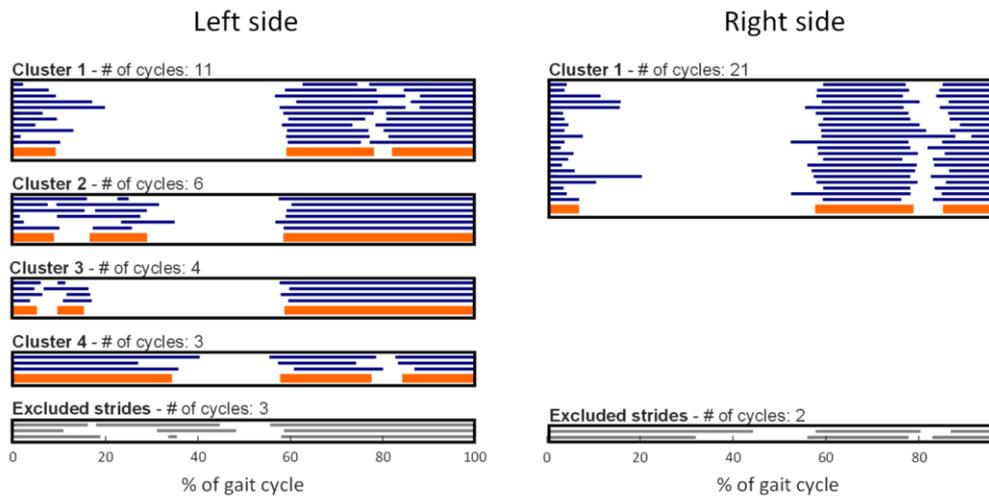
(b) Clustering results – iteration $k-2$

Figure C.2. Clustering results (THA patient, female, 70 years old)

Clustering results obtained cutting the dendrogram at (a) iteration k and (b) $k-2$. As it can be observed from panel (a), the clusters obtained cutting the tree at the iteration k are not an optimal result. In panel (b) the clusters obtained cutting the tree at iteration $k-2$ are represented: in this case, both on the left and right side, two different activation patterns can be recognized, while in panel (a) only one cluster per side is identified.



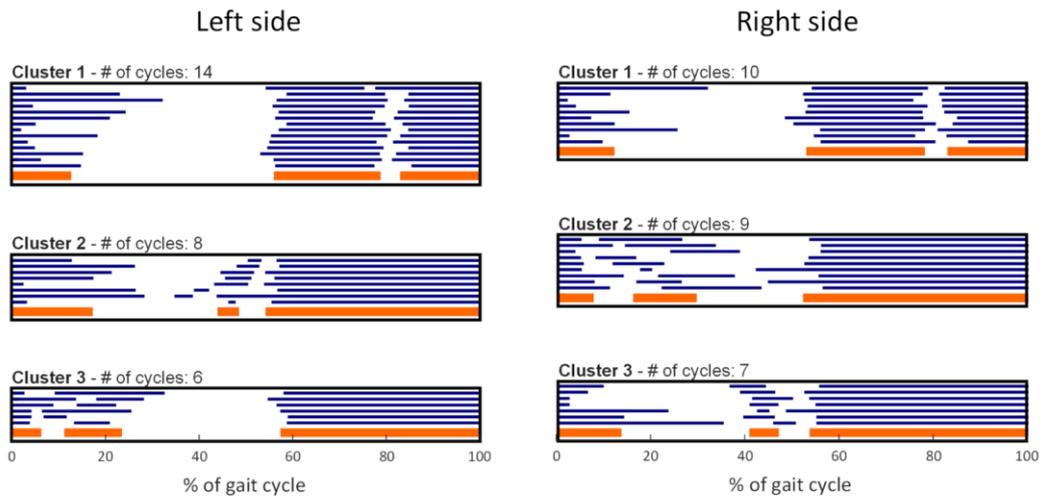
(a) Clustering results – iteration k



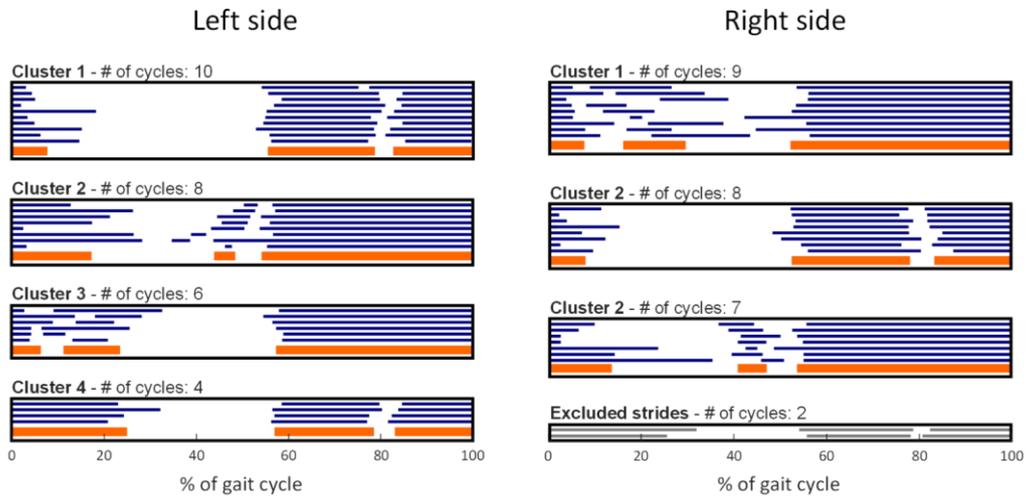
(b) Clustering results – iteration $k-1$

Figure C.3. Clustering results (THA patient, male, 65 years old)

Clustering results obtained cutting the dendrogram at (a) iteration k and (b) $k-1$. As it can be observed from panel (a), the clusters obtained cutting the tree at the iteration k are not an optimal result. In panel (b) the clusters obtained cutting the tree at iteration $k-1$ are represented: in this case, on the left side, is it possible to identified four different activation patterns, while in panel (a) only three clusters are identified.



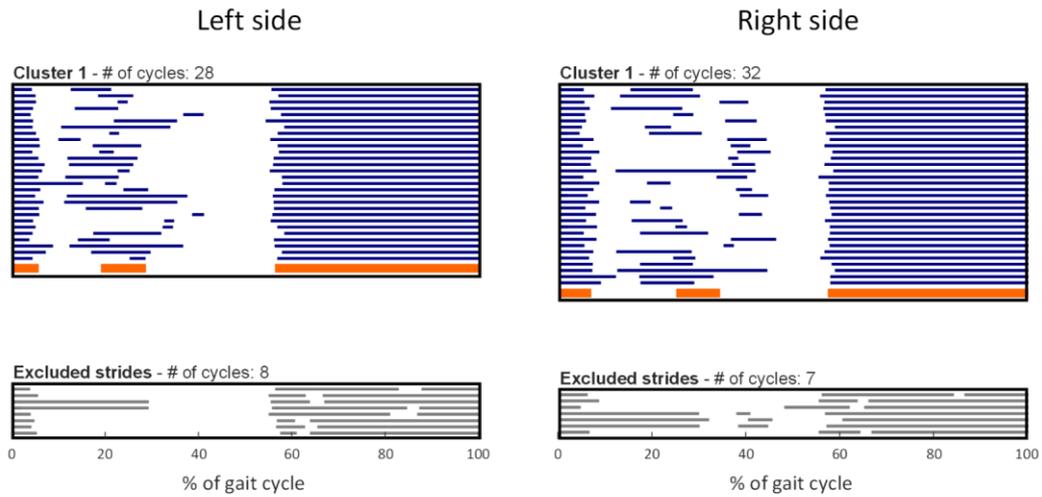
(a) Clustering results – iteration k



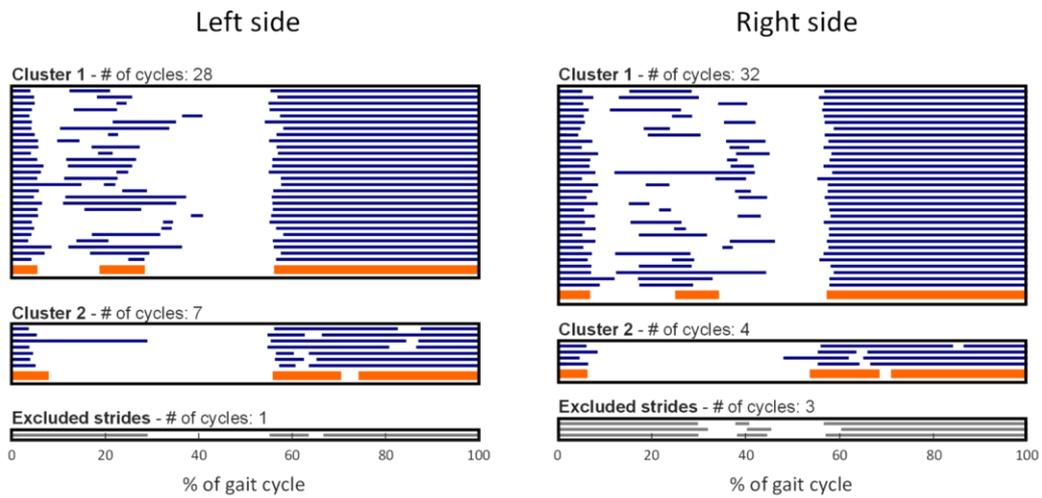
(b) Clustering results – iteration $k-1$

Figure C.4. Clustering results (THA patient, female, 75 years old)

Clustering results obtained cutting the dendrogram at (a) iteration k and (b) $k-1$. As it can be observed from panel (a), the clusters obtained cutting the tree at the iteration k are not an optimal result. In panel (b) the clusters obtained cutting the tree at iteration $k-1$ are represented: in this case, both on the left side, four different activation patterns can be recognized, while in panel (a) only three clusters are identified on the left side.



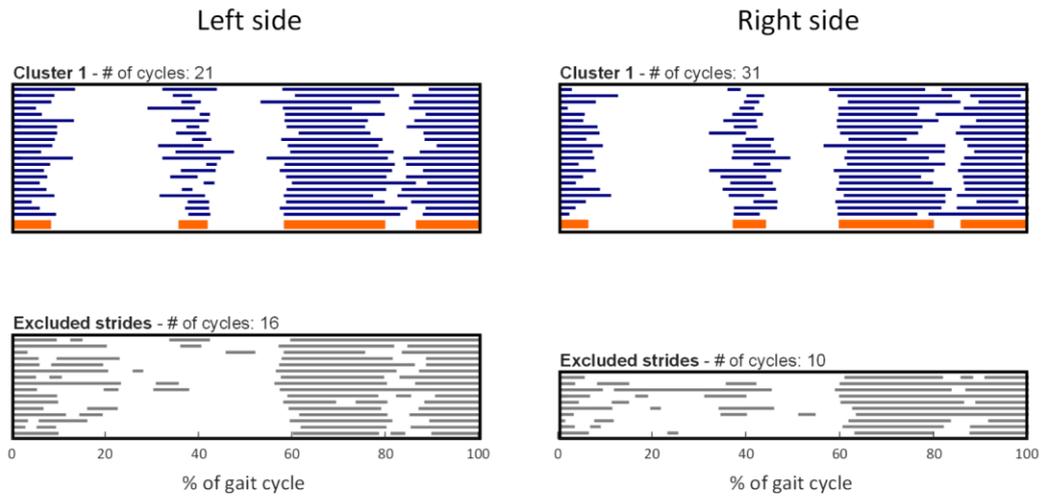
(a) Clustering results – iteration k



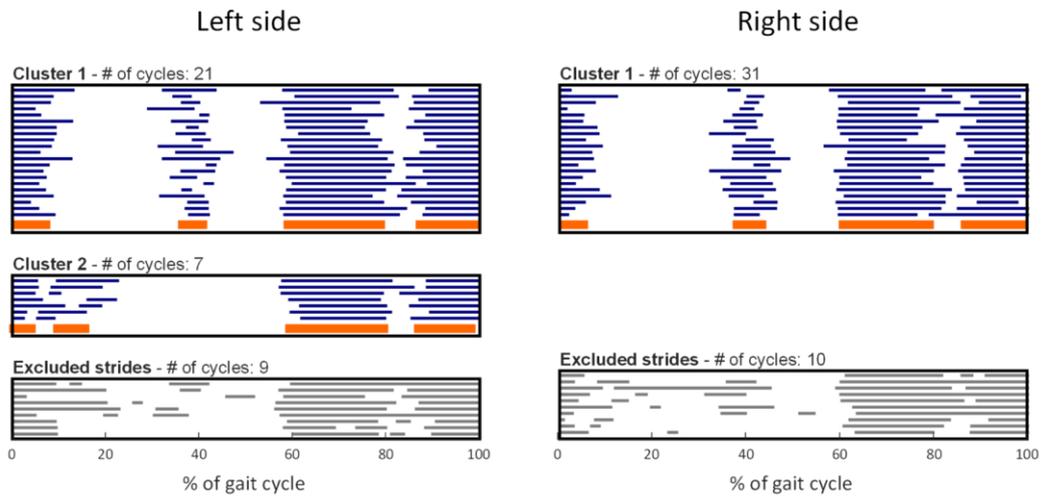
(b) Clustering results – iteration $k+2$

Figure C.5. Clustering results (Healthy male, 51 years old)

Clustering results obtained cutting the dendrogram at (a) iteration k and (b) $k+2$. As it can be observed from panel (a), the clusters obtained cutting the tree at the iteration k are not an optimal result. In panel (b) the clusters obtained cutting the tree at iteration $k+2$ are represented: in this case, both on the left and right side, two different activation patterns can be recognized, while in panel (a) only one cluster per side is identified.



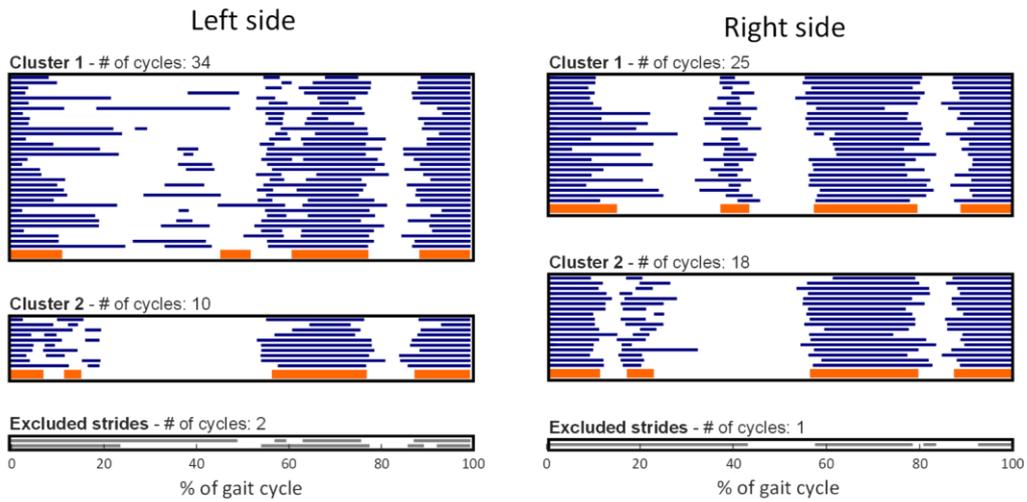
(a) Clustering results – iteration k



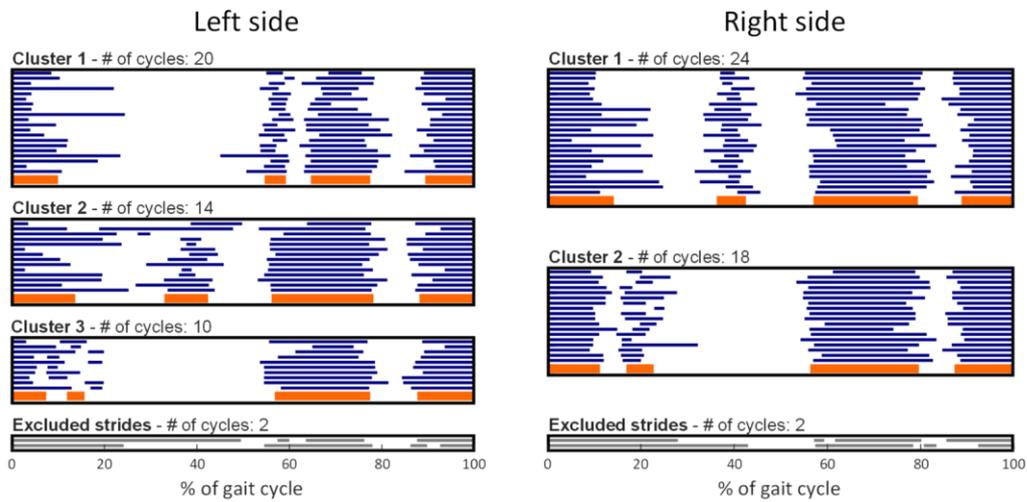
(b) Clustering results – iteration $k+2$

Figure C.6. Clustering results (Healthy male, 24 years old)

Clustering results obtained cutting the dendrogram at (a) iteration k and (b) $k+2$. As it can be observed from panel (a), the clusters obtained cutting the tree at the iteration k are not an optimal result. In panel (b) the clusters obtained cutting the tree at iteration $k+2$ are represented: in this case, on the left side, it is possible to identified two different activation patterns, while in panel (a) only one cluster is identified.



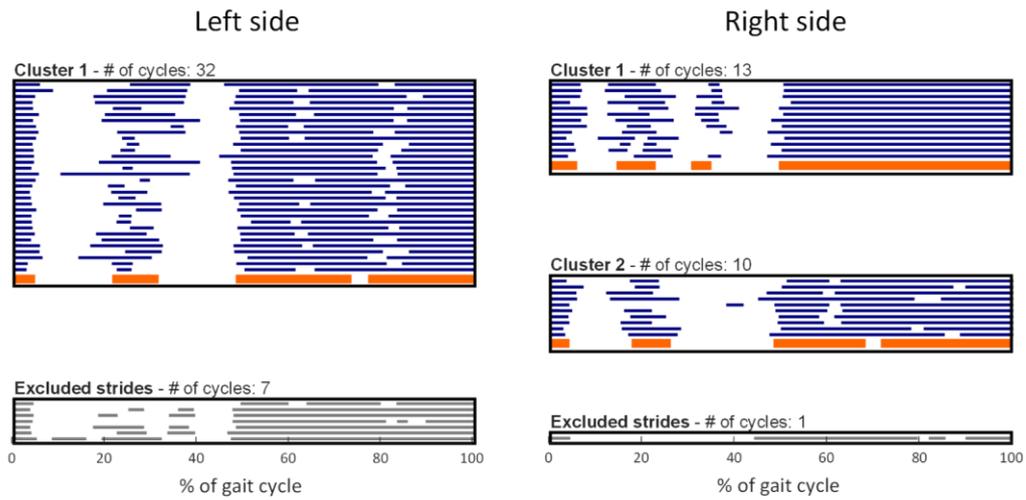
(a) Clustering results – iteration k



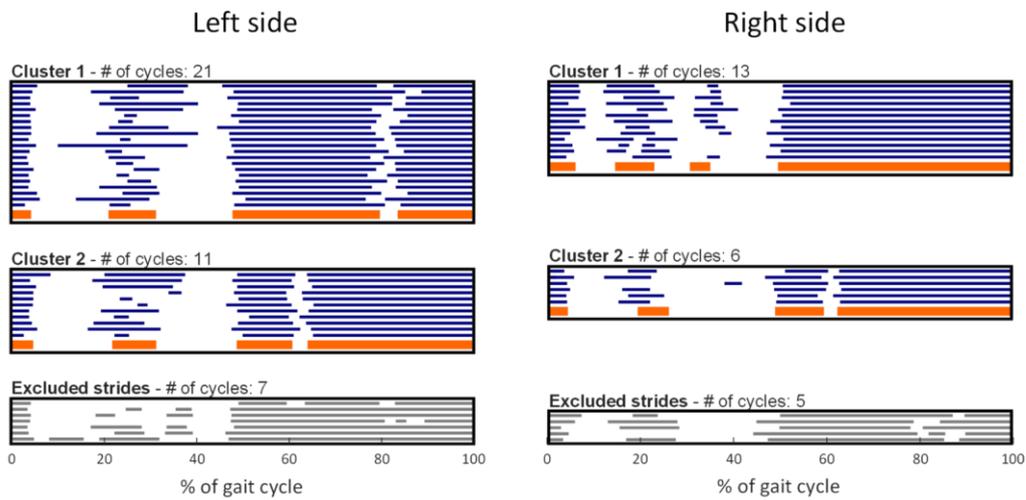
(b) Clustering results – iteration $k-1$

Figure C.7. Clustering results (Healthy female, 28 years old)

Clustering results obtained cutting the dendrogram at (a) iteration k and (b) $k-1$. As it can be observed from panel (a), the clusters obtained cutting the tree at the iteration k are not an optimal result. In panel (b) the clusters obtained cutting the tree at iteration $k-1$ are represented: in this case, on the left side, is it possible to identified three different activation patterns, while in panel (a) only two clusters are identified.



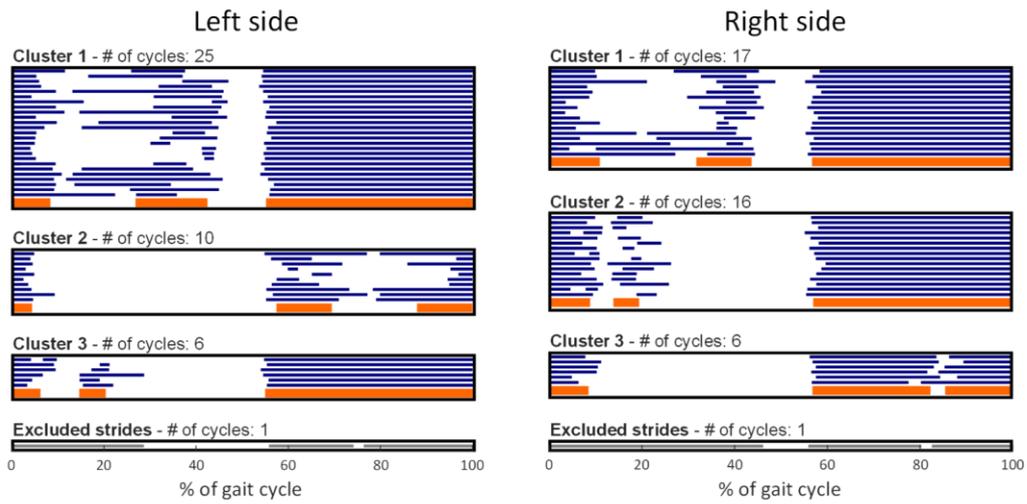
(a) Clustering results – iteration k



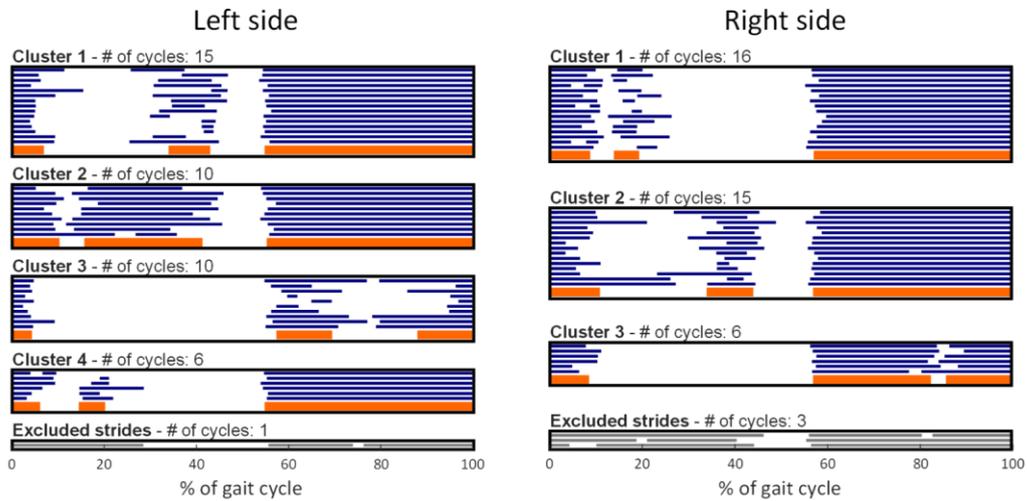
(b) Clustering results – iteration $k-1$

Figure C.8. Clustering results (Healthy male, 30 years old)

Clustering results obtained cutting the dendrogram at (a) iteration k and (b) $k-1$. As it can be observed from panel (a), the clusters obtained cutting the tree at the iteration k are not an optimal result. In panel (b) the clusters obtained cutting the tree at iteration $k-1$ are represented: in this case, on the left side, is it possible to identified two different activation patterns, while in panel (a) only one cluster is identified.



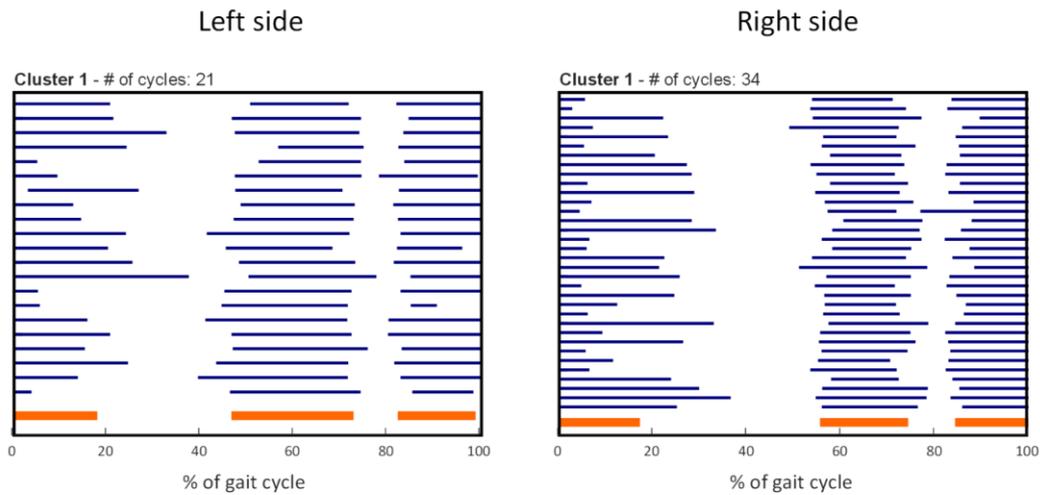
(a) Clustering results – iteration k



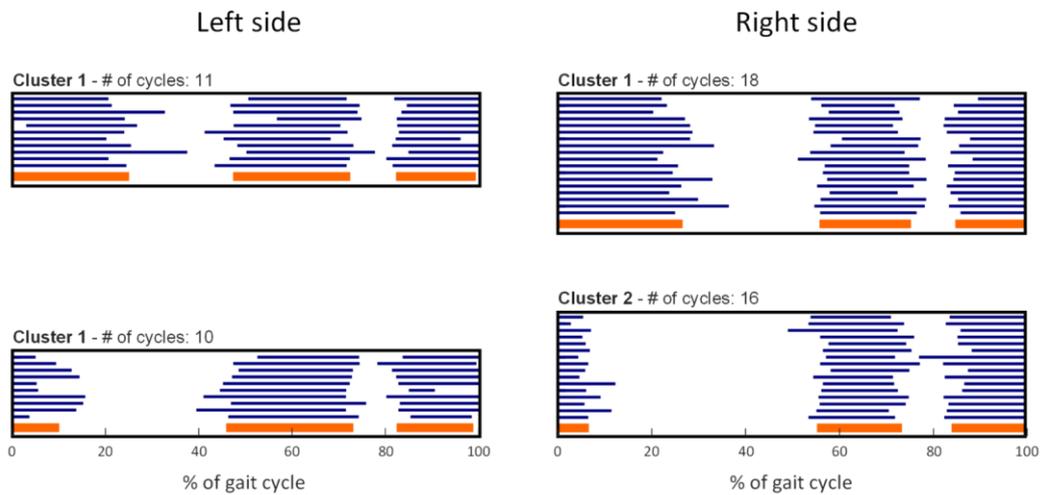
(b) Clustering results – iteration $k-1$

Figure C.9. Clustering results (Healthy female, 25 years old)

Clustering results obtained cutting the dendrogram at (a) iteration k and (b) $k-1$. As it can be observed from panel (a), the clusters obtained cutting the tree at the iteration k are not an optimal result. In panel (b) the clusters obtained cutting the tree at iteration $k-1$ are represented: in this case, on the left side, is it possible to identified four different activation patterns, while in panel (a) only three left clusters are identified.



(a) Clustering results – iteration k



(b) Clustering results – iteration $k-1$

Figure C.10. Clustering results (iNPH patient female, 85 years old)

Clustering results obtained cutting the dendrogram at (a) iteration k and (b) $k-1$. As it can be observed from panel (a), the clusters obtained cutting the tree at the iteration k are not an optimal result. In panel (b) the clusters obtained cutting the tree at iteration $k-1$ are represented: in this case, both on the left and right side, two different activation patterns can be recognized, while in panel (a) only one cluster per side is identified.