POLITECNICO DI TORINO
Repository ISTITUZIONALE

An open-source toolbox for enhancing the assessment of muscle activation patterns during cyclical movements

Original
An open-source toolbox for enhancing the assessment of muscle activation patterns during cyclical movements / Dotti, G., Ghislieri, M., Castagneri, C., Agostini, V., Knaflitz, M., Balestra, G., Rosati, S.. - In: PHYSIOLOGICAL MEASUREMENT. - ISSN 0967-3334. - ELETTRONICO. - 45:(2024). [10.1088/1361-6579/ad814f]

Availability:
This version is available at: 11583/2992944 since: 2024-10-11T07:36:15Z

Publisher:
IOP

Published
DOI:10.1088/1361-6579/ad814f

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IOP postprint/Author's Accepted Manuscript

“This is the accepted manuscript version of an article accepted for publication in PHYSIOLOGICAL MEASUREMENT. IOP Publishing Ltd is not responsible for any errors or omissions in this version of the manuscript or any version derived from it. The Version of Record is available online at http://dx.doi.org/10.1088/1361-6579/ad814f

(Article begins on next page)

06 November 2026

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

ACCEPTED MANUSCRIPT • OPEN ACCESS

An open-source toolbox for enhancing the assessment of muscle activation patterns during cyclical movements

To cite this article before publication: Gregorio Dotti *et al* 2024 *Physiol. Meas.* in press https://doi.org/10.1088/1361-6579/ad814f

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2024 The Author(s). Published on behalf of Institute of Physics and Engineering in Medicine by IOP Publishing Ltd**.

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 4.0 licence, this Accepted Manuscript is available for reuse under a CC BY 4.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by/4.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 5.90.107.30 on 30/09/2024 at 17:29

IOP Publishing Physiological Measurement
Journal **XX** (XXXX) XXXXXX https://doi.org/XXXX/XXXX

An Open-Source Toolbox for Enhancing the Assessment of Muscle Activation Patterns during Cyclical Movements

Gregorio Dotti[1,2,*], Marco Ghislieri[1,2], Cristina Castagneri[1,2], Valentina Agostini[1,2], Marco Knaflitz[1,2], Gabriella Balestra[1,2], and Samanta Rosati[1,2]

[1] BIOLAB, Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy
[2] PoliTo[BIO]Med Lab, Politecnico di Torino, Turin, Italy

E-mail: gregorio.dotti@polito.it

Abstract

Objective. The accurate temporal analysis of muscle activations is of great importance in several research areas spanning from the assessment of altered muscle activation patterns in orthopaedic and neurological patients to the monitoring of their motor rehabilitation. Several studies have highlighted the challenge of understanding and interpreting muscle activation patterns due to the high cycle-by-cycle variability of the sEMG data. This makes it difficult to interpret results and to use sEMG signals in clinical practice. To overcome this limitation, this study aims at presenting a toolbox to help scientists easily characterize and assess muscle activation patterns during cyclical movements. **Approach**. *CIMAP* (Clustering for the Identification of Muscle Activation Patterns) is an open-source Python toolbox based on agglomerative hierarchical clustering that aims at characterizing muscle activation patterns during cyclical movements by grouping movement cycles showing similar muscle activity. **Main results**. From muscle activation intervals to the graphical representation of the agglomerative hierarchical clustering dendrograms, the proposed toolbox offers a complete analysis framework for enabling the assessment of muscle activation patterns. The toolbox can be flexibly modified to comply with the necessities of the scientist. *CIMAP* is addressed to scientists of any programming skill level working in different research areas such as biomedical engineering, robotics, sports, clinics, biomechanics, and neuroscience. CIMAP is freely available on GitHub (https://github.com/Biolab-PoliTO/CIMAP). **Significance**. *CIMAP* toolbox offers scientists a standardized method for analyzing muscle activation patterns during cyclical movements.

Keywords: clustering, cyclical movements, EMG, machine learning, muscle activation patterns.

1. Introduction

Surface electromyography (sEMG) is commonly used, in several research areas, to quantitatively and non-invasively assess dynamic muscle activity in both physiological and pathological conditions. Among the most studied sEMG-derived parameters, the identification of the onset and offset instants of muscle activity plays a fundamental role. The assessment of sEMG activation intervals achieved great interest among researchers in a wide variety of clinical, robotic, and sports applications. In particular, muscle activation intervals are used to assess altered sEMG patterns in patients affected by orthopaedic or neurological diseases (Hsu *et al* 2019, Castagneri *et al* 2019), to define rehabilitation protocols tailored to the patient needs (Akef Khowailed and Abotabl 2019), to study posture control (Labanca *et al* 2021),

Accepted Manuscript

 © xxxx IOP Publishing Ltd

to control prostheses and exoskeletons (Micera *et al* 2010, Li *et al* 2023), and to evaluate return-to-sport of athletes after injury (Rocchi *et al* 2020).

However, sEMG signals during gait are characterized by high cycle-by-cycle variability that makes it difficult to interpret the results and to use sEMG data in clinical practice (Agostini *et al* 2020, Winter and Yack 1987). For a specific muscle of a subject, different activation patterns are usually assessed during cyclical movements, each of them characterized by a specific frequency of occurrence (Di Nardo *et al* 2017). Considering the walking task, even in healthy subjects, a single muscle does not show a single preferred pattern of activation, but up to 4-5 distinct sEMG patterns, each characterized by a different number of activation intervals occurring within the stride (Agostini *et al* 2010, 2015). To overcome this limitation, specific algorithms are needed to help scientists to easily characterize and assess muscle activation patterns during cyclical movements.

Cluster analysis may represent a useful tool for helping scientists to study the different muscle activation patterns during cyclical movements. In this perspective, $CIMAP$ (Clustering for the Identification of the Muscle Activation Patterns) algorithm was proposed and validated in different healthy and pathological conditions (Rosati *et al* 2017a, 2017b). The $CIMAP$ algorithm is based on agglomerative hierarchical clustering and aims at characterizing muscle activation patterns during cyclical movements by grouping movement cycles showing similar muscle activity. This algorithm requires the muscle activation intervals as input data, computed through a muscle activity detector (not included in the toolbox). It was specifically developed to assess muscle activity patterns during walking in both physiological and pathological conditions and it was successfully applied to the study of gait asymmetry in healthy, orthopaedic, and neurological patients (Castagneri *et al* 2019, Rosati *et al* 2021). Moreover, the study by Ghislieri *et al.* used the $CIMAP$ algorithm as a pre-processing step before muscle synergy extraction to evaluate human motor control during locomotion (Ghislieri *et al* 2019, 2020). Notice that the $CIMAP$ algorithm was originally proposed for assessing gait in both physiological and pathological conditions. Nevertheless, $CIMAP$ can be potentially applied to other cyclical movements, as cycling and reach-to-grasp movements.

In this contribution, to support researchers interested in the analysis of muscle activation patterns, we distribute an open-source Python toolbox ($CIMAP$) that allows for obtaining all the representative muscle activation patterns of a muscle. The number of clusters identified by the $CIMAP$ toolbox and the cluster size (i.e., the number of elements belonging to the same cluster), may represent meaningful information in clinics, since they indicate how many sEMG patterns were found and how frequently they occur during the analyzed movement (Agostini *et al* 2014). The proposed toolbox adopts an object-oriented programming approach that allows a clear definition of a few classes incorporating data structure and data processing methods, empowering researchers to easily extend and customize the toolbox to meet specific data and protocol needs. To better describe the processing pipeline and to provide a set of practical guidelines, an example of $CIMAP$ application is presented considering sEMG signals acquired from a lower-limb muscle of a representative healthy subject during a 5-minute walking task.

Researchers with little coding experience will find in the Python toolbox $CIMAP$ a complete framework for the assessment of muscle activation patterns during cyclical movements, from the pre-processing of muscle activation intervals to the graphical representation of the clustering results.

2. Methods

$CIMAP$ toolbox is implemented in Python and includes all the required steps for performing the analysis of muscle activation patterns. This toolbox incorporates functions for all the analysis steps, from data preparation to the graphical representation of the clustering results and data saving.

Figure 1 shows the workflow of the $CIMAP$ algorithm. The following section provides a brief description of the main functions implemented within the $CIMAP$, including references to relevant literature:

a) *Dataset preparation*: $CIMAP$ toolbox requires as input the muscle activation intervals (i.e., the time intervals characterized by muscle activity). Notice that this toolbox does not include a muscle activation interval detection step. Therefore, before using this toolbox, researchers should first apply a muscle activity detector. Input data should be provided in a **.csv* file containing a $M \times (N + 1)$ matrix representing muscle activation intervals, with M being the number of muscles acquired and N being the number of time samples after time-normalization. From the muscle activation intervals, the onset and offset instants are extracted, representing the beginning and the end of each muscle activation, respectively. More details about input data format are freely available on the BIOLAB GitHub repository (https://biolab-polito.github.io/CIMAP/data_requirements.html);

b) *Pre-processing*: movement cycles characterized by the same number of activation intervals occurring within the cycle duration are grouped into sub-groups (or "modalities"). Modalities characterized by a small number of movement cycles ($num.\,cycles < Th$) are considered as non-representative and thus discarded

Accepted Manuscript

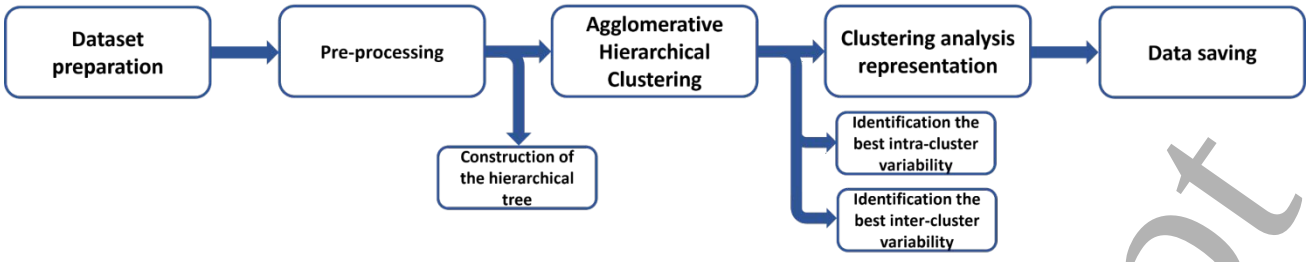

Figure 1: Schematic representation of the steps implemented within $CIMAP$ algorithm.

from the following analyses. Based on a previous study by Dotti *et al.* (Dotti *et al* 2021), the value of Th was set equal to 10 movement cycles. Notice that this value of Th was selected and optimized based on activation intervals acquired from lower-limb muscles during long-lasting walking. More details about the optimization process can be found in the study by Dotti *et al.* (Dotti *et al* 2021). Th value can be easily adjusted to meet specific data needs, such as a small number of movement cycles or different motor tasks;

c) *Agglomerative hierarchical clustering*: considering each modality separately, agglomerative hierarchical clustering is applied. From a number of clusters equal to the number of cycles belonging to the modality under consideration (i.e., each cluster is characterized by a single movement cycle), agglomerative hierarchical clustering iteratively merges the two "closest" clusters, until a single cluster is obtained (i.e., a single cluster containing all the movement cycles belonging to the modality under consideration). The complete linkage method is used to select the two "closest" clusters to be merged, considering both Manhattan (L_1 norm) and Chebyshev (L_∞ norm) as distance metrics (Kaufman and Rousseeuw 1990). Thus, for each modality, agglomerative hierarchical clustering is applied twice (the first time considering L_1 norm and the second time considering L_∞ norm as distance metric). The cutoff points of the dendrograms (i.e., the final number of clusters) are selected applying to each dendrogram the cutoff rule proposed by Rosati *et al.* (Rosati *et al* 2017b). In particular, three cutoff points are identified to evaluate the difference (Δ_{it}) in inter-cluster distances between consecutive clustering iterations. The first cutoff point is identified as the first iteration in which the Δ_{it} is higher than the Δ_{it} average over all the iterations. The second cutoff point is identified as the first iteration in which the Δ_{it} is higher than the Δ_{it} average plus one standard deviation of Δ_{it} over all the iterations. The third cutoff point is identified as the point where the moving average of the Δ_{it} series, starting from the last iteration and moving backward, stops decreasing monotonically. Then, the best cutoff point is chosen as the one showing the lowest value in the index (CUT_IND) defined by Rosati *et al.* (Rosati *et al* 2017b) as follows:

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

where $INTRA_VAR_i$ represents the intra-cluster variability of the i-th cluster calculated as the mean pairwise distance between cycles, n represents the number of meaningful clusters (defined as those with more than one element), and C_i is the number of cycles included in the i-th meaningful cluster. Finally, after comparing the dendrograms obtained considering the L_1 norm and L_∞ norm metrics, the one showing the lowest intra-cluster variability is selected for the following analyses;

d) *Clustering analysis representation*: this toolbox includes several visualization methods that allow for the examination of data throughout the entire analysis process. In particular, dendrograms showing clustering results can be represented for each muscle and each modality, separately. An example of clustering results representation is provided in **Figure 5**;

e) *Data saving*: to increase the accessibility of results, their interpretability and interoperability, clustering results can be exported in *.csv* format. More details about output data format are freely available on the BIOLAB GitHub repository (https://biolab-polito.github.io/CIMAP/data_requirements.html);

Further details about the implemented Python classes and the default setting parameters are freely available on the GitHub repository (https://github.com/Biolab-PoliTO/CIMAP).

3. Results

This section describes all the steps involved in the analysis of muscle activation patterns of a sample dataset of sEMG data acquired from two lower-limb muscles (left and right Lateral

Accepted Manuscript

Gastrocnemius muscle) of a healthy subject during a 5-minute overground walking task.

The first step is the loading of muscle activation intervals contained in the sample dataset ('*input_file*') through the '*data_reading*' function. The '*data_reading*' function can be called as follows:

```
s,muscles = CIMAP.data_reading(input_file =
input_file)
```

where s represents a data structure containing the muscle activation intervals to be processed through the *CIMAP* algorithm and *muscles* represents a data dictionary containing sEMG information (i.e., muscle labels, side, sensor placement, … .).

Then, the '*remove_add_ints*' function is called to remove (if any) outliers of muscle activation intervals (i.e., movement cycles characterized by always-ON or always-OFF muscle activation patterns). Further details about the outlier removal process can be found in the study by Rosati *et al.* (Rosati *et al* 2017b). The outlier removal process can be performed through the '*remove_add_ints*' function as follows:

```
s = CIMAP.remove_add_ints(s)
```

where s contains the muscle activation intervals after the outlier removal step.

Muscle activation intervals can be graphically represented using the '*act_plot*' function. In the following, the '*act_plot*' function is called to represent all the muscle activation intervals of the left and right Lateral Gastrocnemius muscles (LGS_L and LGS_R, respectively):

```
CIMAP.act_plot(s,target='LGS')
```

where s represents the data structure containing the pre-processed muscle activation intervals and *target* is a variable containing the labels of the muscles to be represented.

Figure 2 shows the output of the '*act_plot*' function considering data from the sample dataset included in the toolbox. More specifically, it represents the muscle activation intervals obtained from the left and right LGS muscles of a healthy volunteer during a 5-minute overground walking. Each horizontal blue line represents a muscle activation interval extracted from a single gait cycle expressed in percentage of cycle duration (blue = muscle active, white = muscle inactive). It can be observed that, despite intra-cycle variability, LGS muscle activity mainly occurs between 20% and 50% of the gait cycle for both the left and right sides. The muscle activation intervals included in the sample dataset were computed using the LSTM-MAD algorithm proposed by Ghislieri *et al.* (Ghislieri *et al* 2021).

Before clustering, muscle activation intervals are divided into modalities (i.e., movement cycles characterized by the same number of activation intervals occurring within the cycle duration) by using the '*modality_division*' function as follows:

```
muscles = CIMAP.modality_division(s,muscles)
```

where s contains the muscle activation intervals and *muscles* represents the data dictionary suitable for the following clustering analysis.

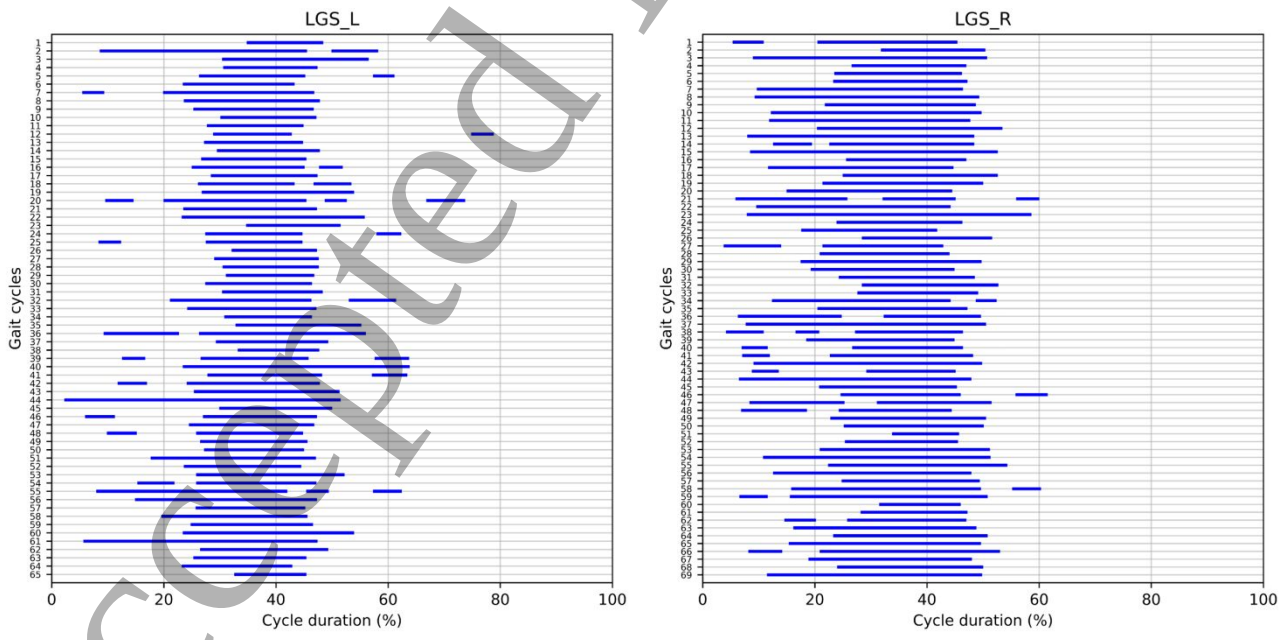

Figure 2: Example of muscle activation intervals for the left and right Lateral Gastrocnemius muscles (LGS_L and LGS_R, respectively) of a healthy volunteer during a 5-minute overground walking task. The blue horizontal lines represent muscle activation intervals expressed as a percentage of the cycle duration (BLUE = muscle active, WHITE = muscle inactive).

Accepted Manuscript

To visualize the modality distribution and to assess the number of movement cycles belonging to each modality, the '*modality_distribution*' function can be used as follows:

```
CIMAP.modality_distribution(s,target='LGS')
```

where *s* contains the muscle activation intervals and *target* is the variable containing the labels of the muscles to be represented.

Figure 3 shows the output of the '*modality_distribution*' function representing the histogram of the movement modalities extracted from the muscle activation intervals of the left and right Lateral Gastrocnemius muscles (LGS_L and LGS_R, respectively) included in the sample dataset. In particular, **Figure 3** shows the number of gait cycles belonging to each modality for both the left and right sides. For example, considering the right side, more than 50 gait cycles are characterized by a single muscle activation (Modality 1), approximately 15 gait cycles are characterized by 2 muscle activations (Modality 2), and less than 5 gait cycles are characterized by 3 muscle activations (Modality 3).

The agglomerative hierarchical clustering is then performed using the '*dendrograms*' function, which computes two different dendrograms by using the L_1 norm and L_∞ norm, respectively. To select the optimal cutting point from the two dendrograms (L_1 norm and L_∞ norm), the '*cuts*' function is applied to each dendrogram. The optimal cutting point is selected based on the rules defined by Rosati *et al.* (Rosati *et al* 2017b). The clustering analysis can be performed by calling the '*dendrograms*' and '*cuts*' functions as follows:

```
muscles = CIMAP.dendrograms(muscles)
muscles = CIMAP.cuts(muscles)
```

The '*dendrograms*' and '*cuts*' functions can be easily customized by users to meet specific data and protocol needs. The toolbox documentation available on GitHub (https://github.com/Biolab-PoliTO/CIMAP) includes further details on the '*dendrograms*' and '*cuts*' functions.

Clustering results can be graphically represented through the '*dendro_plot*' function as follows:

```
CIMAP.dendro_plot(muscles,target='LGS')
```

where *muscles* contains the clustering results obtained through the '*dendrograms*' and '*cuts*' functions and *target* is the variable containing the labels of the muscles to be represented.

Figure 4 represents the output of the '*dendro_plot*' function. For each movement modality, the computed dendrograms are represented. Above each dendrogram, the specific metric and cutting point used are represented, as defined by Rosati *et al.* (Rosati *et al* 2017b). The clusters identified after the cutting point selection are represented in different colors.

To save the CIMAP output, the '*algorithm_output*' and '*result_saver*' functions can be called as follows:

```
cimap_output = CIMAP.algorithm_output(s,muscles)
CIMAP.result_saver(cimap_output)
```

where *cimap_output* is a data dictionary containing the clustering results of each muscle after discarding non-significant modalities.

CIMAP results are saved in an easy-to-read and open-source format (*.csv*). More specifically, results are in a $M \times (C+1)$ matrix, where M represents the number of muscles and C the total number of cycles. Notice that the first column should contain the labels of each muscle as defined in the input file.

Finally, the sEMG activation intervals clustering computed through *CIMAP* can be represented using the '*cluster_plot*' function as follows:

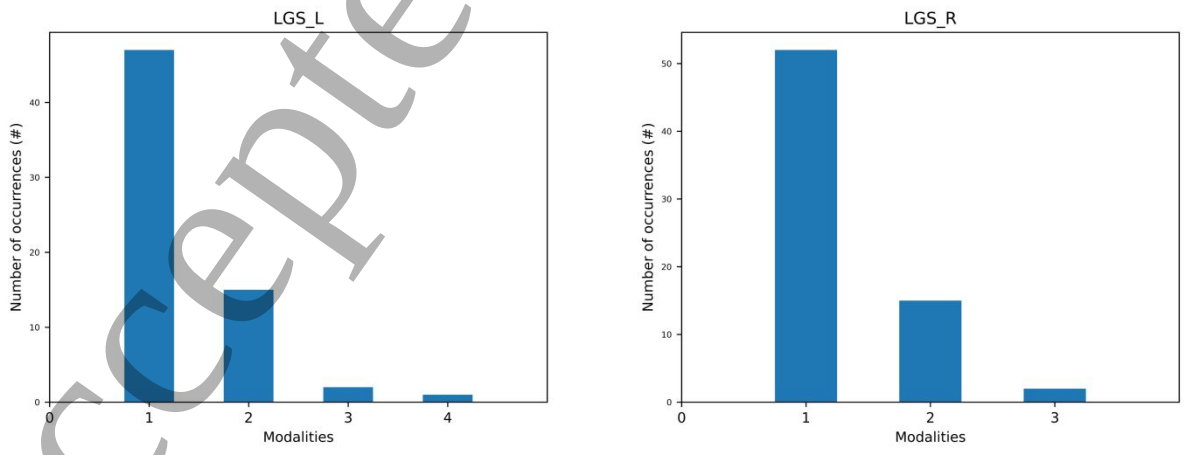

Figure 3: Histograms of movement modalities obtained from the muscle activation intervals of the left and right Lateral Gastrocnemius muscles (LGS_L and LGS_R, respectively) included in the sample dataset. Each histogram represents the number of gait cycles belonging to each modality.

Accepted Manuscript

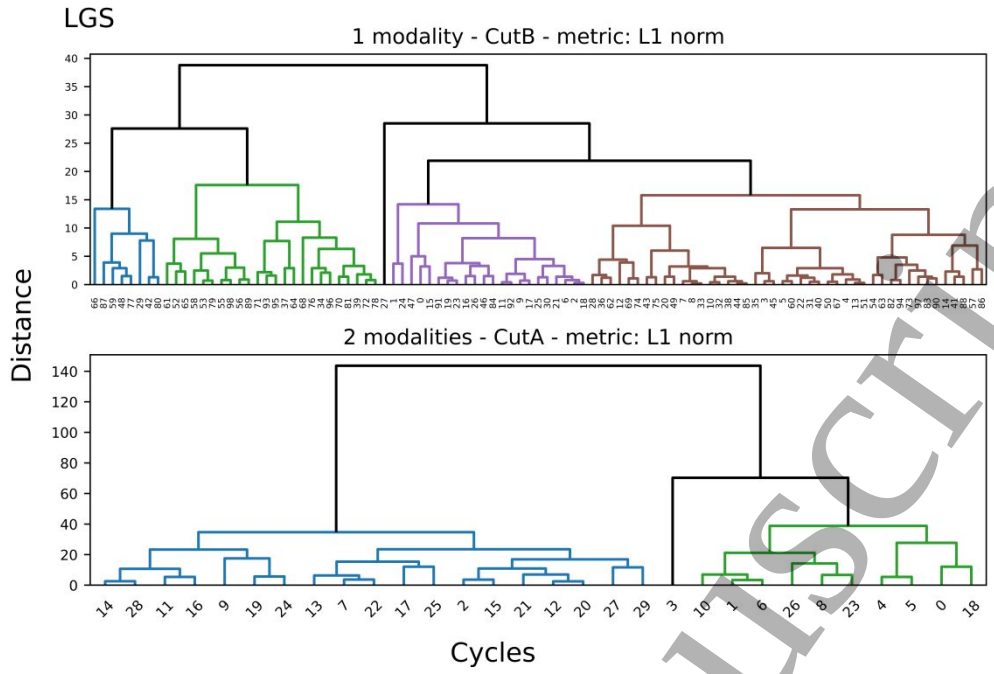

Figure 4: Agglomerative hierarchical clustering dendrograms of the sEMG activation intervals displaying a single modality (at the top) and two modalities (at the bottom). SEMG activation intervals that belong to the same cluster are represented with the same color. Both dendrograms were created using the L_1 norm as the distance metric.

```
CIMAP.clusters_plot(cimap_output,target='LGS',
color = True)
```

where *cimap_output* is a data dictionary containing the clustering results and *target* is the variable containing the labels of the muscles to be represented. The toolbox documentation freely available on GitHub (https://github.com/Biolab-PoliTO/CIMAP) includes further details on the '*cluster_plot*' function.

Figure 5 shows the clustering of the sEMG activation intervals computed through *CIMAP*. The sEMG activation intervals are color-coded according to the colors used in the agglomerative hierarchical clustering dendrograms (**Figure 4**). In each row, the colored lines depict the sEMG activation intervals within a gait cycle. The sEMG activation intervals are grouped into clusters, indicated by different colors. The black lines represent the cluster centroids. Gait cycles belonging to non-representative modalities (i.e., characterized by a small number of gait cycles) are represented in the 'Modality under Th = 10' panel.

4. Discussion and Conclusions

CIMAP is an open-source and comprehensive toolbox for the assessment of muscle activation patterns from surface electromyographic (sEMG) data. The proposed toolbox offers a complete analysis framework for enhancing the assessment of muscle activation patterns from pre-processing of muscle activation intervals to the representation of agglomerative hierarchical clustering dendrograms. *CIMAP* adopts an object-oriented programming approach allowing scientists of any programming skill level to easily extend and customize the toolbox to meet specific data and protocol needs. To better explain the toolbox and offer practical guidance, an example of *CIMAP* application was presented. This example involved analyzing a sample dataset of sEMG signals acquired from a lower-limb muscle of a healthy subject during a 5-minute walk.

CIMAP requires as input the muscle activation intervals extracted from the sEMG data of the muscles of a subject. However, this toolbox does not include a muscle activation interval detection step. Therefore, researchers who want to analyze muscle activation patterns using *CIMAP* should first apply a muscle activity detector before using this toolbox. In the last years, several muscle activity detectors have been proposed, spanning from approaches based on single- (Hodges and Bui 1996, Solnik *et al* 2008) or double-threshold (Bonato *et al* 1998) to more complex approaches based on machine- (Di Nardo *et al* 2022) or deep-learning techniques (Ghislieri *et al* 2021). In particular, muscle activation intervals included in the sample dataset were computed using the LSTM-MAD algorithm proposed by Ghislieri *et al.* (Ghislieri *et al* 2021).

It is well known that muscle activations are characterized by high cycle-to-cycle variability that may strongly reduce the interpretability of the results. In this perspective, *CIMAP*

Accepted Manuscript

represents a first resource to be used for dealing with variability in muscle activation patterns analysis during

Accepted Manuscript

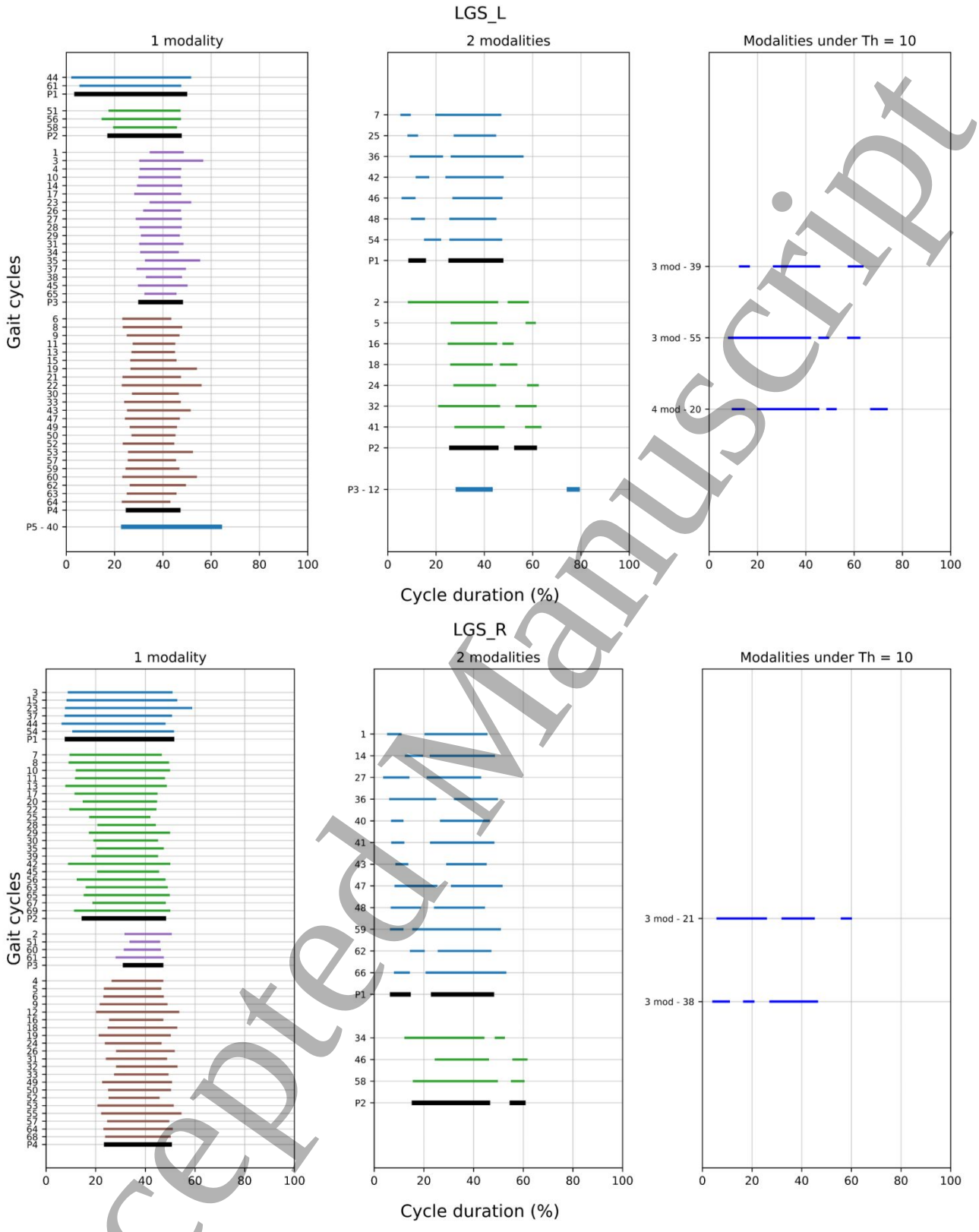

Figure 5: Example of clustering of sEMG activation intervals (muscles: left and right lateral gastrocnemius) computed by means of $CIMAP$. In each row, the colored lines represent the sEMG activation intervals within a gait cycle. SEMG activation intervals are grouped in clusters (represented by different colors) based on CIMAP output. Black lines represent cluster centroids. Gait cycles belonging to non-representative modalities (i.e., characterized by a small number of gait cycles) are represented in the 'Modality under Th = 10' panel.

Accepted Manuscript

cyclical movements. In the last years, *CIMAP* was validated in clinics considering different kinds of disorders affecting gait (such as orthopedic and neurological diseases). In particular, *CIMAP* was used by Castagneri *et al.* (Castagneri *et al* 2019) for the evaluation of gait asymmetry in adults with megaprosthesis of the knee after bone tumor resection, in elderly subjects after total hip arthroplasty, children suffering from hemiplegic cerebral palsy, and in elderly patients affected by idiopathic normal pressure hydrocephalus. Results demonstrated that gait asymmetry, evaluated through *CIMAP*, consistent with the expected impact of the pathologies on the muscle activation during gait, suggesting the applicability of the method for the objective assessment of asymmetry. Moreover, in the study by Rosati *et al.* (Rosati *et al* 2021), *CIMAP* algorithm allows the definition of two quantitative indexes for the assessment of subject muscle coordination, enabling clinicians to identify muscle activation patterns that significantly deviate from those of a reference population.

Even if the *CIMAP* toolbox was originally developed for clinical gait analysis, the clustering approach is independent from the set of muscles considered, it can be easily extended to the study of other cyclical movements, and it can be applied to research areas different from clinics and rehabilitation (e.g., ergonomics, robotics, and sports).

From *CIMAP* outputs, several parameters can be extracted to deeply understand motor control strategies during movement. For example, the size and variability of each representative cluster can be easily extracted to study muscle activation pattern consistency of a subject muscle over the task duration. During walking tasks, healthy subjects are characterized by an increased cluster size (i.e., gait cycles are described by the same number of muscle activation intervals within the gait cycle duration) and a reduced within-cluster variability (i.e., gait cycles are characterized by similar onset and offset timings within gait cycle duration) compared to pathological conditions. Thus, these parameters can be used to distinguish between physiological and pathological gait conditions.

Notice that the set of input parameters implemented in the *CIMAP* toolbox was tested and optimized for clinical gait analysis. In case of different tasks or different numbers of movement repetitions, we recommend testing the method and (if necessary) tuning the input parameters. In particular, the input parameters that may require a tuning step are: (*i*) the threshold Th used to define representative modalities, (*ii*) the minimum number of elements required to define a significant cluster, and (*iii*) the distance metric used for the agglomerative hierarchical clustering.

One of the limitations of this toolbox is the absence of a dedicated Graphical User Interface (GUI). Although a GUI could improve accessibility, custom code may be needed to adapt the assessment of muscle activation intervals for different movements and datasets, which could be challenging to implement in a GUI. However, the *CIMAP* toolbox includes several visualization functions that enable researchers to easily track each processing step. Another limitation of the toolbox is the absence of a muscle activation interval detection step. Nevertheless, the main goal of the toolbox is to offer scientists a standardized method for analyzing muscle activation patterns. This allows researchers the possibility to extract muscle activation intervals based on the specific requirements of their dataset.

In conclusion, an open-source Python toolbox for the assessment of muscle activation intervals was presented to help scientists to easily interpret muscle activation patterns during cyclical movements. This approach might provide a step forward to the understanding of motor control strategies from muscle activation intervals in different pathological conditions affecting movement.

5. Data availability

The latest stable release of *CIMAP* toolbox, the detailed documentation, and a sample dataset are freely available on GitHub (https://github.com/Biolab-PoliTO/CIMAP).

Acknowledgments

The authors acknowledge the support from Politecnico di Torino (Turin, Italy) through the Read&Publish initiative.

CRediT authorship contribution statement

Gregorio Dotti: Data curation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing; **Marco Ghislieri**: Data curation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing; **Cristina Castagneri**: Methodology, Writing – review & editing; **Valentina Agostini**: Conceptualization, Methodology, Writing – review & editing; **Marco Knaflitz**: Conceptualization, Writing – review & editing; **Gabriella Balestra**: Conceptualization, Methodology, Writing – review & editing; **Samanta Rosati**: Conceptualization, Methodology, Writing – review & editing, Supervision.

All the authors have read and agreed to the published version of the manuscript.

References

Agostini V, Lo Fermo F, Massazza G and Knaflitz M 2015 Does texting while walking really affect gait in young adults? *J Neuroeng Rehabil* **12** 1–10

Accepted Manuscript

Agostini V, Ganio D, Facchin K, Cane L, Moreira Carneiro S and Knaflitz M 2014 Gait parameters and muscle activation patterns at 3, 6 and 12 months after total hip arthroplasty *Journal of Arthroplasty* **29** 1265–72

Agostini V, Ghislieri M, Rosati S, Balestra G and Knaflitz M 2020 Surface Electromyography Applied to Gait Analysis: How to Improve Its Impact in Clinics? *Front Neurol* **11** 1–13

Agostini V, Nascimbeni A, Gaffuri A, Imazio P, Benedetti M G and Knaflitz M 2010 Normative EMG activation patterns of school-age children during gait *Gait Posture* **32** 285–9

Akef Khowailed I and Abotabl A 2019 Neural muscle activation detection: A deep learning approach using surface electromyography *J Biomech* **95** 109322

Bonato P, D'Alessio T and Knaflitz M 1998 A statistical method for the measurement of muscle activation intervals from surface myoelectric signal during gait *IEEE Trans Biomed Eng* **45** 287–99

Castagneri C, Agostini V, Rosati S, Balestra G and Knaflitz M 2019 Asymmetry Index in Muscle Activations *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27** 772–9

Dotti G, Ghislieri M, Rosati S, Agostini V, Knaflitz M and Balestra G 2021 The Effect of Number of Gait Cycles on Principal Activation Extraction *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* 985–8

Ghislieri M, Agostini V and Knaflitz M 2019 How to Improve Robustness in Muscle Synergy Extraction *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* vol 2019 (IEEE) pp 1525–8

Ghislieri M, Agostini V and Knaflitz M 2020 Muscle Synergies Extracted Using Principal Activations: Improvement of Robustness and Interpretability *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **28** 453–60

Ghislieri M, Cerone G L, Knaflitz M and Agostini V 2021 Long short-term memory (LSTM) recurrent neural network for muscle activity detection *J Neuroeng Rehabil* **18** 1–15

Hodges P W and Bui B H 1996 A comparison of computer-based methods for the determination of onset of muscle contraction using electromyography *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control* **101** 511–9

Hsu W C, Chang C C, Lin Y J, Yang F C, Lin L F and Chou K N 2019 Assessment on muscle contraction sequences in post-stroke patients during sit-to-stand *Sensors (Switzerland)* **19**

Kaufman L and Rousseeuw P J 1990 Finding Groups in Data

Labanca L, Mosca M, Ghislieri M, Agostini V, Knaflitz M and Benedetti M G 2021 *Muscle activations during functional tasks in individuals with chronic ankle instability: a systematic review of electromyographical studies* vol 90 (Elsevier B.V.)

Li H, Guo S, Bu D, Wang H and Kawanishi M 2023 Subject-Independent Estimation of Continuous Movements Using CNN-LSTM for a Home-Based Upper Limb Rehabilitation System *IEEE Robot Autom Lett* **8** 6403–10

Micera S, Carpaneto J and Raspopovic S 2010 Control of hand prostheses using peripheral information *IEEE Rev Biomed Eng* **3** 48–68

Di Nardo F, Mengarelli A, Strazza A, Agostini V, Knaflitz M, Burattini L and Fioretti S 2017 A new parameter for quantifying the variability of surface electromyographic signals during gait: The occurrence frequency *Journal of Electromyography and Kinesiology* **36** 25–33

Di Nardo F, Nocera A, Cucchiarelli A, Fioretti S and Morbidoni C 2022 Machine Learning for Detection of Muscular Activity from Surface EMG Signals *Sensors 2022, Vol. 22, Page 3393* **22** 3393 Online: https://www.mdpi.com/1424-8220/22/9/3393/htm

Accepted Manuscript

Rocchi J E, Labanca L, Laudani L, Minganti C, Mariani P P and Macaluso A 2020 Timing of muscle activation is altered during single-leg landing tasks after anterior cruciate ligament reconstruction at the time of return to sport *Clinical Journal of Sport Medicine* **30** Online: https://journals.lww.com/cjsportsmed/fulltext/2020/11000/timing_of_muscle_activation_is_altered_during.17.aspx

Rosati S, Agostini V, Knaflitz M and Balestra G 2017a Muscle activation patterns during gait: A hierarchical clustering analysis *Biomed Signal Process Control* **31** 463–9

Rosati S, Castagneri C, Agostini V, Knaflitz M and Balestra G 2017b Muscle contractions in cyclic movements: Optimization of CIMAP algorithm *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* 58–61

Rosati S, Ghislieri M, Dotti G, Fortunato D, Agostini V, Knaflitz M and Balestra G 2021 Evaluation of Muscle Function by Means of a Muscle-Specific and a Global Index

Solnik S, DeVita P, Rider P, Long B and Hortobágyi T 2008 Teager-Kaiser operator improves the accuracy of EMG onset detection independent of signal-to-noise ratio *Acta Bioeng Biomech* **10** 65–8

Winter D A A and Yack H J J 1987 EMG profiles during normal human walking: stride-to-stride and inter-subject variability *Electroencephalogr Clin Neurophysiol* **67** 402–11

Accepted Manuscript