*#Title*: CIMAP: Clustering for Identification of Muscle Activation Patterns

*#Tags*: clustering, dendrogram, EMG, locomotion, machine-learning algorithms, muscle activity, python

*#Authors*: Gregorio Dotti, Marco Ghislieri, Samanta Rosati, and Gabriella Balestra

*#Affiliations*: Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy; PoliToBIOMed Lab, Politecnico di Torino, Turin, Italy

*#Summary*

Dynamic muscle activity can be quantitatively and non-invasively investigated during several cyclical movements by acquiring surface electromyographic (sEMG) signals. The accurate temporal analysis of muscle activations is of great interest in several research areas, spanning from the assessment of altered muscle activation patterns in orthopaedic or neurological patients to the monitoring of their motor rehabilitation. However, due to the high intra-cycle variability of the muscle activation patterns [], specific algorithms are needed to help scientists to easily characterize and assess muscle activation patterns during cyclical patterns. In this perspective, Clustering for the Identification of Muscle Activation Patterns (```CIMAP```) [] is a Python [] algorithm based on agglomerative hierarchical clustering that aims at characterizing muscle activation patterns during cyclical movements by grouping cycles showing a similar muscle activity. From muscle activation intervals to the graphical representation of the clustering results, the proposed algorithm offers a complete analysis framework for assessing muscle activation patterns that can be flexibly modified at need and applied to cyclical movements different from walking. ```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.

*#Statement of Need*

Surface electromyography (sEMG) is commonly used, in several research areas, to qualitatively and non-invasively assess dynamic muscle activity in both physiological and pathological conditions. Among the sEMG-based analyses, the temporal analysis of muscle activations achieved great interest in several research areas, such as the assessment of altered muscle activation patterns in patients affected by orthopaedic or neurological diseases and motor rehabilitation monitoring []. Muscle activation intervals can be obtained from raw sEMG signals by detecting the beginning (*onset*) and the end (*offset*) time-instant of muscle activations during specific movements. In the last years, several algorithms have been published in the literature, spanning from approaches based on single-threshold [] or double-threshold statistical detectors [] to more complex approaches based on deep-learning techniques []. However, independently from the accuracy of the detector in identifying the onset-offset timing patterns, sEMG signals are characterized by high cycle-by-cycle variability that may reduce muscle activity pattern interpretability. For example, literature reports that during walking, even in healthy subjects, a single subject’s 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 (called *modalities*) []. In this perspective, cluster analysis represents a helpful tool for helping scientists to study the various activation patterns of muscles during a cyclical task. More specifically, a systematic approach for grouping together cycles with homogeneous onset-offset timing patterns can be obtained through cluster analysis. Thus, Clustering for the Identification of the Muscle Activation Patterns (```CIMAP```) algorithm [] was recently proposed in the literature. The ```CIMAP``` algorithm 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 []. Nevertheless, it can be potentially applied to other cyclical movements (e.g., reach-to-grasp movement). Starting from muscle activation intervals, this algorithm allows for obtaining all of the representative activation patterns of a subject’s muscle, each corresponding to a cluster’s prototype. Moreover, the number of clusters identified 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 [], respectively. Notice that, even if the ```CIMAP``` algorithm was originally developed for gait analysis, the clustering approach is independent from the muscle (or set of muscles) considered, can be easily extended to the study of other sEMG cyclical signals, and can be applied not only in the rehabilitation setting, but also in human movement science, ergonomics, and sport. Researchers with little to none coding experience will find in the Python algorithm ```CIMAP``` a complete framework for the assessment of muscle activation patterns during cyclical movements, from muscle activation intervals pre-processing to the graphical representation of the clustering results.

*#Analysis workflow*

The typical workflow when using the ```CIMAP``` algorithm consists of the following five main steps:

1. Data preparation (i.e., to read input data and convert them into the format needed for the following steps);
2. Data pre-processing (i.e., to split input data based on the number of muscle activations within each cycle);
3. Agglomerative hierarchical clustering []:
   * Cutting point identification based on the intra-cluster variability [];
   * Distance metric selection based on the lowest inter-cluster variability [];
4. Clusters' representation (available also at points 3 and 4, see **Figure 1** for a representative example);
5. Data saving (clustering results are saved in an easy-to-read and open-source format).

A typical muscle activation interval analysis can be synthetically written as follows:

>>>s,muscles = CIMAP.data\_reading(input\_file = input\_file) # Load input data

>>>s = CIMAP.removeaddints(s) # Remove atypical activation intervals

>>>muscles = CIMAP.modalitydivision(s,muscles) # Cycles division by modality

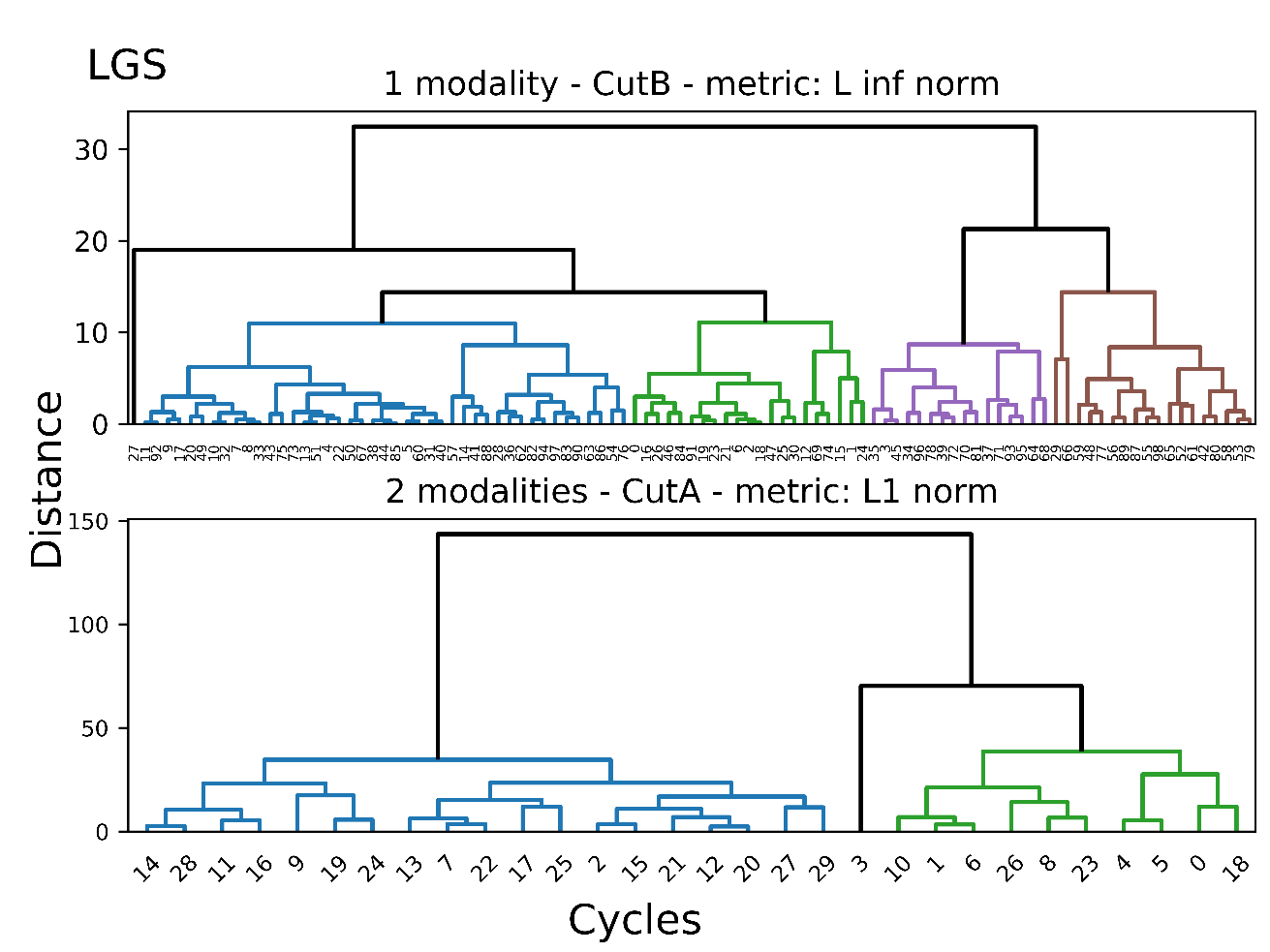
>>>muscles = CIMAP.dendrograms(muscles) # Apply hierarchical clustering

>>>muscles = CIMAP.cuts(muscles) # Define dendrogram cutoff point

>>>cimap\_out = CIMAP.algorithm\_output(s,muscles) # Create output dictionary

>>>CIMAP.resultsaver(cimap\_out) # Save CIMAP results

Default setting parameters are specifically optimized for the analysis of human locomotion. Further details are available on [GitHub]( https://marcoghislieri.github.io/CIMAP/CIMAP.html).



Dendrograms of hierarchical cluster analysis performed on cycles showing a single activation interval (top) and on cycles showing two different activation intervals (bottom), separately. Clusters obtained after the selection of the optimal cutoff point are represented in different colours. Swere extracted from the Lateral Gastrocnemius (LGS) muscle ofThis representation was generated using ```CIMAP``` v1.0.0.

*#Availability*

The latest stable release of ```CIMAP``` is freely available on [GitHub]( https://github.com/marcoghislieri/CIMAP). Documentation and a representative complete example are freely available in each version’s readme file. The latest stable release of ```CIMAP``` can be easily installed through the bash shell or command prompt with the following command:

```pip install CIMAP```

Further details on the setup process and Python minimum requirements are available on [GitHub]( https://github.com/marcoghislieri/CIMAP).

*#Acknoledgments*

The authors are grateful to (in alphabetical order): Prof. Valentina Agostini, Dr. Cristina Castagneri, and Prof. Marco Knaflitz for their contributions to the development and validation of this approach.

*#References*