Muscle Synergies Extracted Using Principal Activations: Improvement of Robustness and Interpretability

(Article begins on next page)

Muscle Synergies Extracted Using Principal Activations: Improvement of Robustness and Interpretability

Marco Ghislieri, Student Member, IEEE, Valentina Agostini, Member, IEEE, and Marco Knaflitz, Member, IEEE

***Abstract*— The muscle synergy theory has been widely used to assess the modular organization of the central nervous system (CNS) during human locomotion. The pre-processing approach applied to the surface electromyographic (sEMG) signals influences the extraction of muscle synergies. The aim of this contribution is to assess the improvements in muscle synergy extraction obtained by using an innovative pre-processing approach. We evaluate the improvement in terms of the possible variation in the number of muscle synergies, of the intra-subject consistency, of the robustness, and of the interpretability of the results. The pre-processing approach presented in this paper is based on the extraction of the muscle principal activations (muscle activations strictly necessary to accomplish a specific biomechanical task) from the original sEMG signals, to then obtain muscle synergies using principal activations only. The results herein presented show that the application of this novel approach for the extraction of the muscle synergies provides a more robust and easily interpretable description of the modular organization of the CNS with respect to the standard pre-processing approach.**

***Index Terms*—Electromyography, EMG, gait analysis, interpretability, muscle activations, muscle synergies, robustness, principal activations.**

I. Introduction

In the last years, the muscle synergy concept was proposed for the assessment of the modular organization of the central nervous system (CNS) during different motor tasks. According to this theory, the CNS controls a small number of muscles rather than coordinating every single muscle involved in a specific motor task. The effectiveness of the muscle synergies in modeling the complexity of the motor control during movements has been demonstrated in several studies and research fields such as clinics, robotics, and sports [1]–[3].

Human locomotion is a complex motor task due to the different biomechanical functions carried out during each gait cycle [4], the multiple degrees of freedom of the skeletal muscle system, and the high intra-cycle variability of the muscle activation patterns [5]. Previous studies demonstrated that human locomotion can be modeled by a small set of muscle synergies responsible of specific biomechanical functions [6]–[9]. The number of muscle synergies extracted and their biomechanical functions may vary depending on the type and the number of muscles acquired [10]. On the average, five muscle synergies are needed to properly describe human locomotion [6], [11].

Muscle synergies during gait are often extracted from surface electromyography (sEMG) signals by applying the Non-Negative Matrix Factorization (NNMF) algorithm [12], [13]. Before applying the factorization algorithm, the acquired sEMG signals are pre-processed to obtain an accurate description of the modular organization of the CNS. In the last years, several pre-processing schemes were proposed in literature for the extraction of muscle synergies during cyclic movements [3], [6], [10], [12]–[15]. As an example, in the work by Clark *et al.* [15] the acquired sEMG signals were previously high-pass filtered with a cut-off frequency of 40 Hz by means of a zero lag fourth-order digital Butterworth filter, demeaned, rectified and smoothed with a zero-lag 4th-order Butterworth digital filter with a cut-off frequency of 4 Hz. Finally, to facilitate the comparison among different muscles and different motor conditions, the filtered sEMG signals were amplitude-normalized with respect to the global maximum of the signal generated by each observed muscle. In the work of Steele et al. [10] sEMG data were band-pass filtered with a lower cut-off frequency of 20 Hz and a higher cut-off frequency of 400 Hz, rectified and low-pass filtered at 10 Hz.

The application of these "standard" techniques may not be sufficient to fully understand the motor control mechanism due to the high intra-cycle variability of the sEMG activation intervals [16]. Statistical Gait Analysis (SGA) [17] was recently introduced to deal with this issue by selecting only the representative gait patterns. It was successfully applied to the study of the frequency-of-occurrence of muscle activation modalities [16], muscle activation timing [18] and co-contractions [19]. Furthermore, through the application of the CIMAP algorithm [20], [21], it is possible to define the principal and secondary muscle activations. Principal activations (PAs) are those activations that are strictly necessary to accomplish the motor task, while the secondary activations (SAs) have an auxiliary function, such as providing

M. Ghislieri, V. Agostini, and M. Knaflitz are with the Department of Electronics and Telecommunications of Politecnico di Torino and with PoliToBIOMedLab of Politecnico of Torino, 10129 Torino, Italy. (e-mail: marco.ghislieri@polito.it, valentina.agostini@polito.it, and marco.knaflitz@polito.it).

corrections to motion and body segment posture. The CIMAP algorithm was successfully applied to the study of gait asymmetry of healthy, orthopedic, and neurological patients [22]–[24].

The aim of this work is to assess how the application of innovative pre-processing techniques (CIMAP with the extraction of PAs) can be used to overcome the limitations of the standard pre-processing algorithms in terms of intra-subject consistency, robustness, and interpretability of muscle synergies.

II. Materials and Methods

A. Sample Population and Experimental Protocol

A sample population of 22 healthy subjects (age: 39.2 years ± 17.0 years, gender: 18 females and 4 males, height: 165.2 cm ± 8.2 cm, weight: 60.9 kg ± 17.5 kg) was retrospectively analyzed using gait data in our database. None of the subjects reported lower limb injuries or had neurological or musculoskeletal disorders that could compromise their gait performance. Twenty subjects out of 22 were right-limb dominant, while two were left-limb dominant, according to the preferred foot to start walking.

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

The experimental protocol conformed to the Helsinki declaration on medical research involving human subjects.

B. Data Acquisitions

Gait data were recorded by means of a multichannel system specifically developed for clinical gait analysis (STEP32, Medical Technology, Italy). The following signals were simultaneously recorded: a) surface electromyographic (sEMG) signals, by means of active probes (configuration: single differential, size: 19 mm × 17 mm × 7 mm, Ag-disks diameter: 4 mm, interelectrode distance: 12 mm, gain: variable in the range from 60 dB to 86 dB); b) foot-switch signals (size: 10 mm × 10 mm × 0.5 mm, activation force: 3 N) to detect gait phases; c) knee joint kinematics signals in the sagittal plane by means of electrogoniometers (accuracy: 0.5°).

The sEMG active probes were positioned over the following 12 muscles of the dominant lower limb and trunk: 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 positioned beneath the heel, the first, and fifth metatarsal heads, bilaterally. Electrogoniometers were positioned on the lateral aspect of the knee joint, bilaterally.

Signals were acquired with a sampling frequency of 2000 Hz, converted by a 12-bit analog to digital converter and sent to a PC for real-time representation. The acquired signals were then imported into the MATLAB® release 2018b (The MathWorks Inc., Natick, MA, USA) to be processed by means of custom routines.

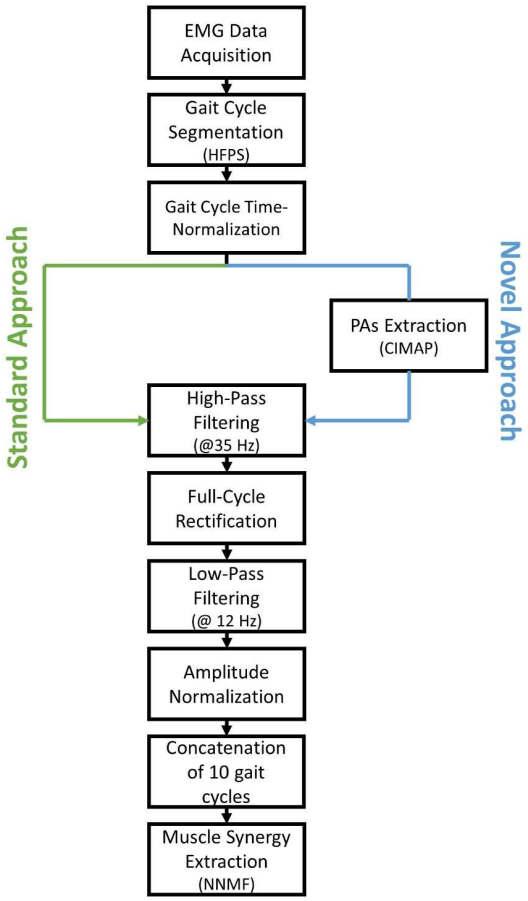

Fig. 1. Schematic description of the two approaches implemented for the muscle synergy extraction: Standard Approach (green) and Novel Approach (blue). In the Novel Approach, only sEMG Principal Activations (PAs) are extracted and considered for the muscle synergy extraction, while in the Standard Approach all the time-samples of the sEMG signals are considered.

C. Data Processing

Before muscle synergy extraction, the acquired sEMG signals were pre-processed using two different approaches: a) the standard approach, in which all the time samples of the sEMG signal were considered, and b) the novel approach in which only the PA intervals of the sEMG signal were considered. Figure 1 shows the workflow of the two approaches implemented for the extraction of the muscle synergies.

1) Gait Cycle Segmentation and Normalization

First, the foot-switch signals were used to segment the gait cycles, keeping only the gait cycles with a normal sequence of gait phases (Heel contact (H), Flat foot contact (F), Push off (P) and Swing (S)), while discarding the atypical gait cycles [25]. Only the gait cycles belonging to the rectilinear path were analyzed, removing those corresponding to direction changes at the beginning and at the end of the walkway (including deceleration before and acceleration after the U-turn) [11]. Then, each segmented gait cycle was time-normalized to 1000 samples [15] and all gait cycles selected were concatenated in a single vector [26].

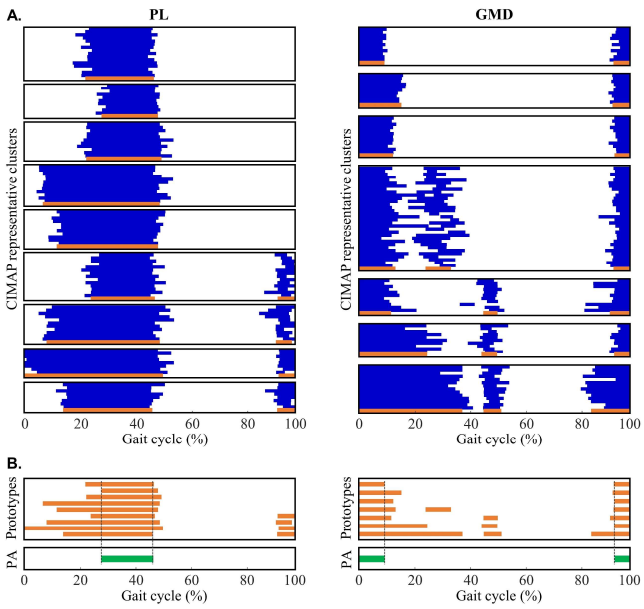

Fig. 2. Example of application of the CIMAP algorithm, on a representative subject, to the Peroneus Longus (PL) and Gluteus Medius (GMD) muscles. (A) Blue intervals inside a rectangle box represent the sEMG activation intervals of the various gait cycles belonging to a cluster. Orange intervals represent the clusters' prototypes. (B) Principal Activations (PAs) are represented in green, defined as the intersection of the clusters' prototypes.

2) *Extraction of Principal Activations (PAs) through CIMAP algorithm*

First, the muscle activation onset/offset intervals were computed from the sEMG signals by means of a double-threshold statistical detector, specifically developed for gait analysis [27]. Then, the optimized version of the CIMAP algorithm (Clustering for Identification of Muscle Activation Patterns) [21] was used to select PAs.

PAs are those muscle activations that are necessary to accomplish a specific biomechanical task and describe the fundamental activation intervals of a specific muscle. The CIMAP algorithm, based on hierarchical clustering, groups together the gait cycles sharing similar sEMG onset-offset activation patterns. For each cluster, the cluster prototype is defined as the median timing pattern. Then, PAs are extracted from the representative clusters, computed as the intersection of the clusters' prototypes [21].

Fig. 2 represents an example of application of the CIMAP algorithm to sEMG gait data acquired during the walking task of a healthy subject from PL and GMD muscles. Figure 2A represents normalized activation intervals for the various gait cycles, grouped in clusters sharing similar activation timings. Each orange interval represents the prototype of a representative cluster. Figure 2B depicts how PAs are obtained from the intersection of the cluster prototypes.

For each gait cycle, the extraction of the PA intervals from the original sEMG signals was performed windowing the time-normalized gait cycle by means of a binary mask that was set to 1 in correspondence of a muscle activation and to 0 if no muscle activation was present.

3) *Muscle Synergy Extraction and Sorting*

The sEMG signals were high-pass filtered through an 8th order Butterworth digital filter with a cut-off frequency of 35 Hz, to remove motion artifact, demeaned, and full-cycle rectified to obtain a non-negative signal. Then, the rectified signals were low-pass filtered by means of a 5th order digital Butterworth filter with a cut-off frequency of 12 Hz to obtain the sEMG envelope. The signals were then normalized in amplitude with respect to the global maximum of each muscle.

The normalized envelopes were divided into groups of 10 concatenated gait cycles (i.e., subgroup 1 contains gait cycles from 1 to 10, subgroups 2 contains gait cycles from 11 to 20, etc.) allowing for muscle synergy assessment over the entire walk [28].

For each subgroup, muscle synergies were extracted from the filtered sEMG signals by means of the Non-Negative Matrix Factorization (NNMF) algorithm. 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) [29] as described in (1):

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

where N represents the optimal number of muscle synergies needed to describe the motor task. The weight vector W_k describes the contribution of each observed muscle to the k-synergy, the activation coefficient vector $C(t)_k$ represents the time-dependent modulation of the muscles recruited in the k-synergy and e represents the prediction error of the factorization algorithm. The MATLAB® function "*nnmf*" was used to apply the NNMF algorithm, setting the routine's input parameters as detailed in TABLE I.

The first algorithm initialization was performed differently for the weight vector W_k and the activation coefficient vector $C(t)_k$. The C matrix was initialized with values randomly selected from a uniform distribution in the range [0, 1]. To improve the performances of the factorization algorithm and the accuracy in the reconstruction of the original sEMG signals, a sparseness constraint was imposed in the initialization of the W matrix [30]. In particular, W matrix was firstly initialized with values randomly chosen from a uniform distribution in the range [0, 0.05], then one random element of each W_k vector was set to a value selected from a uniform distribution in the range [0.7, 0.8]. Therefore, only one muscle for each k-synergy has a

TABLE I
SETTINGS OF THE MATLAB® ROUTINE "*NNMF*" USED FOR MUSCLE SYNERGY EXTRACTION

Settings	**NNMF routine**
Algorithm	multiplicative update
Function tolerance	1e^{-6}
Search tolerance	1e^{-6}
Factorization replicates	50
Factorization iterations (max.)	1000

significant contribution, thus obtaining an extremely sparse initialization [30].

To explore the different solutions of the NNMF algorithm, the factorization process was run many times on the same gait data, changing the number of muscle synergies (*N*) between 1 and 8. For each value of *N*, the reconstruction quality was computed by means of the total Variance Accounted For ($tVAF$), defined as the uncentered Pearson's correlation coefficient expressed in percentage (2):

$$tVAF = \left(1 - \frac{\sum_{k=1}^{m}(M_k - M_k^R)^2}{\sum_{k=1}^{m} M_k^2}\right) \cdot 100 \qquad (2)$$

where *m* represents the number of muscles observed and M_k and M_k^R describes the original and the reconstructed sEMG envelopes of the *k*-synergy, respectively. The optimal number of muscle synergies needed to properly model the sEMG signals of the *i-th* subgroup ($N_{90,i}$) was selected by choosing the smallest number of synergies (*N*) granting a $tVAF$ value equal or greater than 90% [15]. Since each subgroup could be described by a different number of muscle synergies, the final number of synergies (N_{90}) was selected as the mode of the numbers of muscle synergies computed on each 10-gait-cycle subgroup ($N_{90,i}$).

To represent the synergy output in the range [0, 1], the weight vectors W_k were normalized in amplitude with respect to their global maximum. Then, the activation coefficient vectors $C(t)_k$ were multiplied by the correspondent normalized values.

To sort the muscle synergies in the same order for each subgroup, the *k*-means algorithm was applied to the W matrix [9]. Clustering was performed by means of the MATLAB® routine "*kmeans*" using the following input parameters: N_{90} as number of *k*-means clusters, 1000 as maximum number of iterations, 15 as number of replicates and cosine as distance metric. The activation coefficients matrix C was ordered consequently.

D. Muscle Synergy Analysis

The muscle synergies extracted using the standard approach and the novel approach (using PAs), were quantitatively compared in terms of (a) number of muscle synergies, (b) intra-subject consistency, (c) robustness, and (d) interpretability of the results.

a) Number of Muscle Synergies

As stated above, the number of muscle synergies needed to properly reconstruct the sEMG signals of each 10-gait-cycle subgroup was selected by choosing the smallest number of synergies (*N*) granting a $tVAF$ value equal or greater than 90%. The final number of synergies (N_{90}) was then selected as the mode of the number of muscle synergies computed on each 10-gait-cycle subgroup.

b) Intra-Subject Consistency

The intra-subject consistency of the muscle synergies among the subgroups of 10 gait cycles was evaluated by computing the similarity of the previously sorted weight vectors W_k and activation coefficient vectors $C(t)_k$, separately. The similarity between each couple of vectors was assessed by means of the cosine similarity (*CS*) [31]. The *CS* between the vectors of the *i*- and *j*-th subgroup of the *k*-synergy was defined as the normalized scalar product between the vectors expressed in percentage, as described in (3) and (4):

$$CS_{W,k}^{i,j} = \left(\frac{W_k^i \cdot W_k^j}{\|W_k^i\| \, \|W_k^j\|}\right) \cdot 100 \qquad (3)$$

$$CS_{C,k}^{i,j} = \left(\frac{C_k^i \cdot C_k^j}{\|C_k^i\| \, \|C_k^j\|}\right) \cdot 100 \qquad (4)$$

where $CS_{W,k}^{i,j}$ and $CS_{C,k}^{i,j}$ represent the cosine similarity computed between the weight vectors W_k and the activation coefficients $C(t)_k$ of the *i*- and *j-th* subgroup, respectively. The CS values range between 0 (no similarity) and 1 (complete similarity).

c) Robustness

The robustness of the muscle synergies among different subgroups of 10 gait cycles was assessed through the Cross-Variance Accounted For [26] defined as follows:

$$CrossVAF^{i,j} = \left(1 - \frac{\sum_{k=1}^{m}(M_k^i - M_k^{R,j})^2}{\sum_{k=1}^{m}(M_k^i)^2}\right) \cdot 100 \qquad (5)$$

where M_k^i and $M_k^{R,j}$ describe the original and the reconstructed sEMG signals of the *k*-muscle for the *i*- and *j-th* subgroup, respectively. This parameter assesses how well the muscle synergies extracted for the *i-th* subgroup reconstruct the sEMG signals that belong to the *j-th* subgroup. For each subject, the average $CrossVAF$ value was computed over all the possible couples of 10-gait-cycle subgroups. The average $CrossVAF$ can assume values ranging from 0% to 100%, corresponding to low or high correlation, respectively, between the reconstructed and the original sEMG signals belonging to different subgroups.

d) Interpretability

According to the muscle synergy theory, specific

TABLE II
BIOMECHANICAL FUNCTIONS OF THE MUSCLE SYNERGIES DURING GAIT.

Function	Involved muscles	Biomechanical function
F1	TFL, GMD	Hip joint stabilization during heel strike and load acceptance phase.
F2	LGS, PL, SOL	Propulsion at the mid and terminal stance.
F3	TA	Forefoot clearance control during the swing phase and foot control during the first rocker.
F4	MH, LH	Leg deceleration at the end of the swing phase.
F5	LD_R, LD_R	Control of the trunk position in the frontal plane at heel strike.

biomechanical functions can be associated to each muscle synergy. These biomechanical functions were generally assigned to each muscle synergy by observing the prevailing muscles contribution (W_k> 0.5) and the profile of the activation coefficients $C(t)_k$ [4], [5], [11], [32]. Rimini et al. [11] discovered 5 biomechanical function common to all the analyzed subjects during the walking task. TABLE II reports the description of each biomechanical function and the muscles involved in each function. For example, the biomechanical function F2 is used to generate the propulsion and requires, among the observed muscles, the involvement of the LGS, PL and, SOL muscles. All the other muscles are not directly involved in the propulsive function.

The muscle synergy interpretability (I) was then evaluated considering the average of the muscle synergy weights that are not directly involved in the biomechanical function described by the k-synergy. The muscle synergy interpretability was computed as described in (6):

$$I = \frac{1}{N}\sum_{k=1}^{N}((1-\overline{W_k'})\cdot 100) \tag{6}$$

where N represents the optimal number of muscle synergies needed to describe the motor task and $\overline{W_k'}$ is the average contribution of the muscles not involved in the biomechanical task. A muscle synergy can be considered more easily

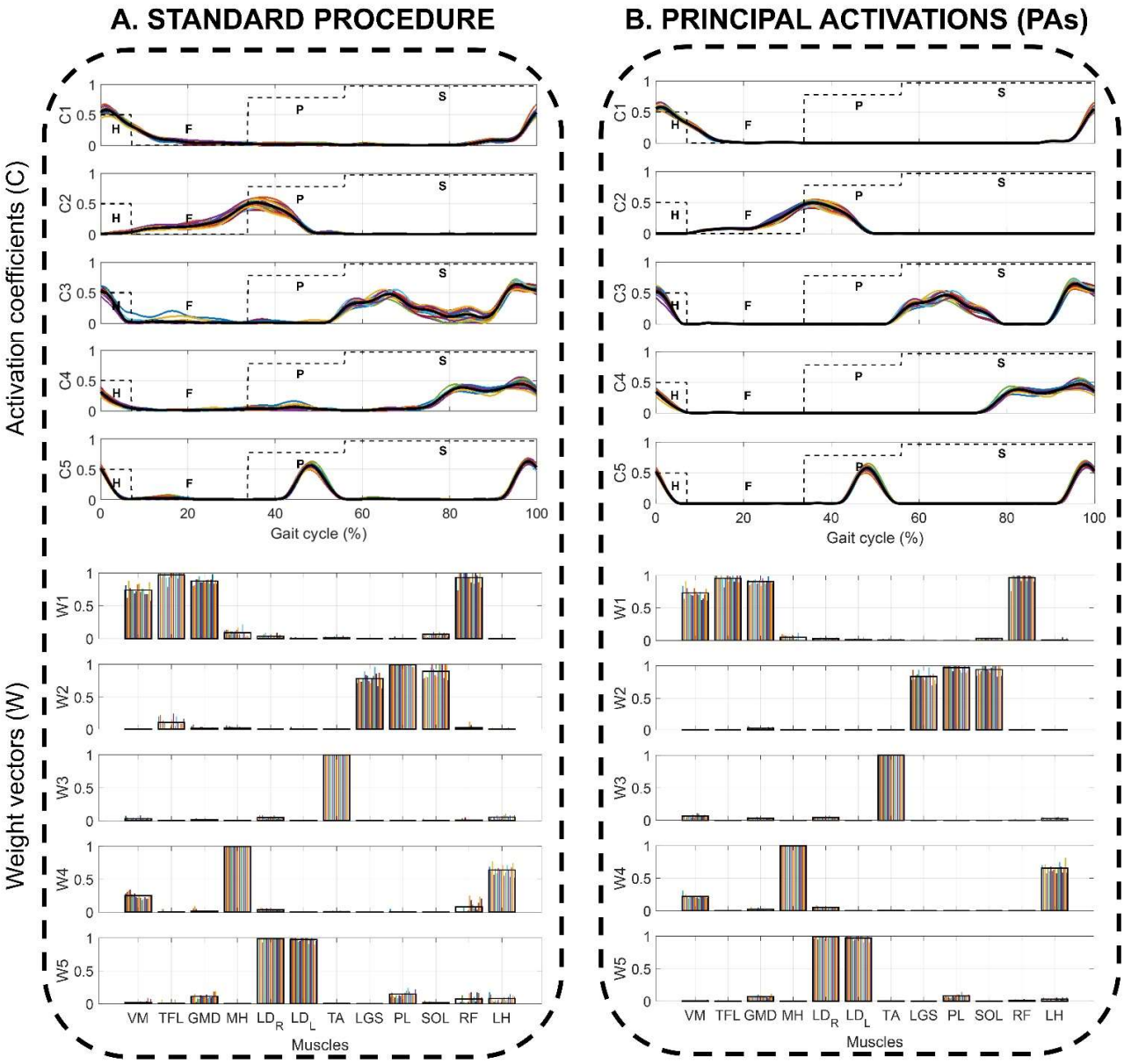

Figure 3. Activation coefficients C_k and weight vectors W_k obtained with two different processing techniques: (A) standard approach, and (B) novel approach with extraction of principal activations (PAs). Each colored line (or colored vertical bar) represents C_k (or W_k) extracted from a single subgroup 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 foot-switch signal with the indication of the 4 gait phases: Heel contact (H), Flat foot contact (F), Push off (P) and Swing (S).

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).

E. Statistical Analysis

The hypothesis of normality of the distribution of the computed parameters was tested by means of the Lilliefors test setting the significance level (α) at 0.05. If the normality hypothesis was rejected, the Wilcoxon signed-rank test was used to compare the results obtained with the two approaches, otherwise the Student *t*-test was implemented (α=0.05).

III. Results

The subjects walked at an average speed of 1.2 m/s ± 0.1 m/s. On the average, a dataset of 156 ± 25 typical gait cycles was analyzed for each subject, divided into 16 ± 2 subgroups of 10 gait cycles each.

In the following, we compare the results obtained with the standard and the novel approach (entailing the extraction of PAs) in terms of (A) number of muscle synergies, (B) intra-subject consistency, (C) robustness, and (D) interpretability.

A. Number of Muscle Synergies

Both pre-processing techniques required the same number of muscle synergies (N_{90}) to properly reconstruct the original sEMG signals with a $tVAF$ higher than 90%. For each of the 22 subjects, 5 muscle synergies were needed to accurately reconstruct the original sEMG signals. Considering the standard approach, the muscle synergies were extracted with an average $tVAF$ value of 93.1% ± 1.5%, while considering the novel approach the muscle synergies were extracted with an average $tVAF$ value of 91.6% ± 1.8%.

Figure 3 reports as an example the muscle synergies extracted from a representative subject of the sample population using the two processing techniques: (Fig. 3A) standard approach and (Fig. 3B) the novel approach (with PAs). The muscle synergies extracted through the standard and novel techniques revealed no significant changes both in the number of muscle synergies and in their composition.

B. Intra-Subject Consistency

No statistically significant differences in intra-subject consistency were found comparing the two approaches (standard and with PAs extraction), for both the weight vectors W_k and the activation coefficients $C(t)_k$. More specifically, both approaches revealed high values of *CS*, suggesting a high similarity among the muscle synergies obtained from different 10-gait-cycle subgroups.

For the weight vectors W_k, *CS* was equal to 0.980 ± 0.002, using the standard approach, and 0.983 ± 0.003 using the novel approach, respectively. The application of the Student's paired *t*-test revealed no significant differences between these approaches (p = 0.27).

For the activation coefficients $C(t)_k$, *CS* was equal to 0.978 ± 0.003 for the standard approach and 0.983 ± 0.004 for the novel approach, respectively. The Student's paired *t*-test revealed no statistically significant differences between these procedures (p = 0.36). Figure 4A shows the values of the intra-subject consistency values computed using the two approaches with the superimposition of the standard errors.

C. Robustness

The robustness assessment revealed a slightly significant increase in the $CrossVAF$ value computed applying the novel approach with respect to the standard approach (Student's paired *t*-test, p = 0.048). In particular, the $CrossVAF$ value was equal to 81.2% ± 0.7%, for the standard approach, and 81.8% ± 0.6%, for the novel approach, respectively. Figure 4B represents the values of the $CrossVAF$ values computed considering the two processing approaches with the superimposition of the standard errors.

D. Interpretability

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's paired *t*-test, p = 0.0007). The computed *I*-value was equal to 90.45% ± 0.52%, for the standard approach, and 92.72% ± 0.62%, for the novel approach, respectively. Figure

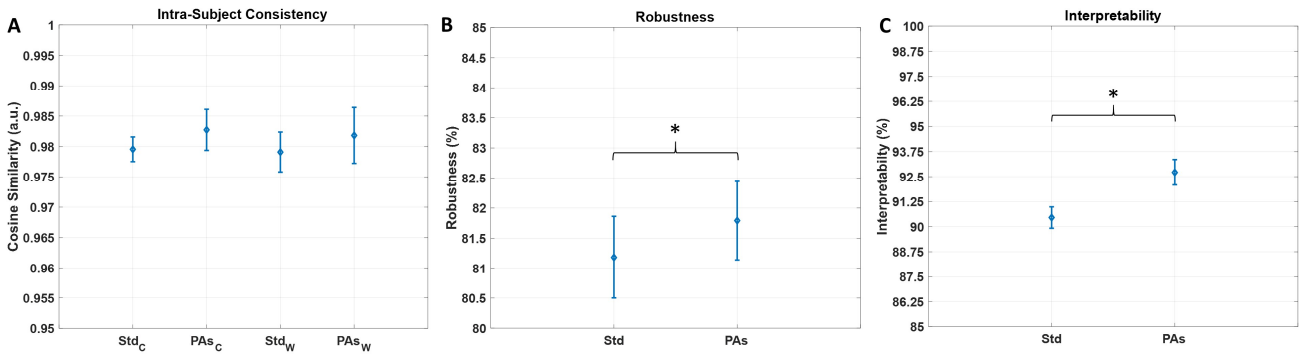

Figure 4. Average (± SE) values of the three parameters computed to compare the standard approach (Standard) and the novel approach (PAs): (A) represents the values of the intra-subject consistency of the weight vectors and the activation coefficients across different subgroup of 10-gait-cycles, (B) shows the vaules of the muscle synergy robustness and (C) the values of the muscle synergy interpretability.

4C shows the values of the muscle synergy interpretability for both the pre-processing techniques with the superimposition of the standard errors.

IV. Discussion

The relevance of the analysis of the muscle synergies is well known in literature and several studies supporting the importance of this theory have been published in the last years [2], [3], [11], [15], [31], [33]. The approaches previously used in literature ("standard" approach) to extract the muscle synergies generally considered the whole sEMG signals as input of the factorization algorithm [6], [10], [13], [15]. These standard approaches may be influenced by the high cycle-by-cycle variability of the sEMG activation patterns [16]. Therefore, these approaches may not be sufficient to fully understand the motor control strategies during human locomotion.

The novel approach presented in this paper can be used to overcome the drawback of the standard techniques by selecting the PA intervals from the original sEMG signals. Considering PAs allow evaluating only the "necessary" muscle activations, discarding those with auxiliary function [22], such as those providing corrections to cyclic motion and body segment posture. According to this approach, only the time-samples in correspondence of PAs are considered as inputs of the factorization algorithm for the extraction of the muscle synergies [11], [21], while the remaining time-samples are set to zero.

To assess the performance of the novel approach with respect to the standard one, the muscle synergies extracted from 22 healthy subjects during a walking task have been compared in terms of number of muscle synergies, intra-subject consistency, robustness, and muscle synergy interpretability.

In terms of the number of muscle synergies, both the approaches accurately reconstruct the original sEMG signals by means of five muscle synergies. These muscle synergies are 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. In particular, the same five biomechanical functions described in Ref. [11] can be associated to the muscle synergies extracted considering each of the two approaches.

In terms of intra-subject consistency, the standard and the novel approach reveal similar values of intra-subject consistency (0.980 ± 0.002 and 0.983 ± 0.003, respectively), suggesting a high repeatability of the motor control strategy among the 10-gait-cycle subgroups, independently from the pre-processing used.

In terms of robustness and interpretability of the muscle synergies, the novel approach outperforms the standard one. Indeed, our results suggest that the extraction of the PAs allows for obtaining a slightly higher robustness ($81.8\% \pm 0.8\%$ vs. $81.2\% \pm 0.7\%$), and a better interpretability ($92.72\% \pm 0.62\%$ vs. $90.45\% \pm 0.52\%$) of the muscle synergies with respect to the standard approach, thus providing a more robust and clear assessment of the modular organization of the CNS during gait.

By analyzing the number and the composition of the muscle synergies and the intra-subject consistency, we found no loss of information due to the extraction of the PAs, with respect to the standard approach. Moreover, the higher performances in terms of robustness and muscle synergy interpretability obtained considering the novel approach demonstrate that the extraction of the PAs may successfully improve the muscle synergy analysis during gait in healthy subjects.

In this study, we proposed the application of an innovative pre-processing technique to extract muscle synergies (CIMAP to select PAs). However, there are also other elements of the processing chain giving raise to the muscle synergies extraction that must be considered. In general, prior literature has drawn considerable attention to the influence of EMG pre-processing on muscle synergy extraction [34]–[36]. In particular, the importance of standardization in envelope estimation [37] and amplitude normalization [38], [39] were thoroughly investigated.

To properly obtain PAs, sEMG signals must be recorded for at least 3 minutes during gait. In this study, we acquired 5 minutes of gait signals, for each subject. Signal recording during a "long" physiological walk is needed to be able to collect at least 100 - 200 valid gait cycles required for PA analysis (in particular for the application of the CIMAP algorithm). Notice that this does not limit the feasibility and applicability of the methodology to pathological populations. Indeed, gait analysis is commonly used to quantitatively assess patients' locomotion performance only in those patients able to independently walk, for some minutes, without walking aids or external support. In the past, several studies demonstrated the feasibility of gait data acquisition, during recording sessions lasting 3 minutes, in patients suffering from different neurological conditions, e.g. Normal Pressure Hydrocephalus [40], mild ataxia [41] and cerebral palsy [42].

In this work, we used the CIMAP algorithm to select PAs and discard secondary activations, suggesting that muscle synergies are better understood when considering only PAs. This is in line with a previous work in which the selection of PAs was used to define a robust asymmetry index, based on EMG activity during locomotion [23]. In particular, the study of Ref. [23] provided a validation both on healthy and pathological populations. The approach proposed in this work was validated considering a population of healthy subjects, but the use of PAs for extracting muscle synergies can be extended to the analysis of subjects affected by neurological disorders, for whom the assessment of motor control through muscle synergies may be of the uttermost importance [43], [44]. Although the validation of the current work has been performed on healthy individuals, the most obvious final goal of future works is providing such a validation also on patients affected by neurological disorders. However, when considering pathological populations, also secondary activations might be fundamental in the interpretation of the results. Indeed, discarding the effect of auxiliary functions included in the secondary activations might bias the interpretation of pathological types of behavior, instead of improving it. Nevertheless, it should be noticed that the opportunity to separate principal from secondary activations, provided by the CIMAP algorithm, does not preclude studying

also secondary (auxiliary) activations. Therefore, it is advisable that future developments of the proposed approach will include a validation also on pathological populations, possibly analyzing both principal and secondary activations.

I. Conclusions

In conclusion, the results presented in this paper demonstrate that the extraction of the principal activations can be successfully used as pre-processing step before the muscle synergy extraction, allowing a more robust and interpretable assessment of the modular organization of the CNS during a walking task without any loss of information. A 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.

References

[1] J. Taborri *et al.*, "Feasibility of Muscle Synergy Outcomes in Clinics, Robotics, and Sports: A Systematic Review," *Appl. Bionics Biomech.*, vol. 2018, pp. 1–19, 2018.

[2] V. C. K. Cheung, A. d'Avella, M. C. Tresch, and E. Bizzi, "Central and Sensory Contributions to the Activation and Organization of Muscle Synergies during Natural Motor Behaviors," *J. Neurosci.*, vol. 25, no. 27, pp. 6419–6434, Jul. 2005.

[3] S. A. Chvatal and L. H. Ting, "Common muscle synergies for balance and walking," *Front. Comput. Neurosci.*, vol. 7, p. 48, May 2013.

[4] J. Perry, "Gait Analysis: Normal and Pathological Function," *J. Pediatr. Orthop.*, vol. 12, p. 815, 1992.

[5] D. A. Winter and H. J. Yack, "EMG profiles during normal human walking: stride-to-stride and inter-subject variability," *Electroencephalogr. Clin. Neurophysiol.*, vol. 67, no. 5, pp. 402–411, Nov. 1987.

[6] Y. P. Ivanenko, R. E. Poppele, and F. Lacquaniti, "Five basic muscle activation patterns account for muscle activity during human locomotion," *J. Physiol.*, vol. 556, no. Pt 1, pp. 267–282, 2004.

[7] V. Monaco, A. Ghionzoli, and S. Micera, "Age-related modifications of muscle synergies and spinal cord activity during locomotion.," *J Neurophysiol*, vol. 104, no. 4, pp. 2092–2102, 2010.

[8] S. A. Chvatal and L. H. Ting, "Voluntary and Reactive Recruitment of Locomotor Muscle Synergies during Perturbed Walking," *J Neurosci*, vol. 32, no. 35, pp. 12237–12250, 2012.

[9] K. M. Steele, M. C. Tresch, and E. J. Perreault, "Consequences of biomechanically constrained tasks in the design and interpretation of synergy analyses.," *J. Neurophysiol.*, vol. 113, no. January, p. jn.00769.2013, 2015.

[10] K. M. Steele, M. C. Tresch, and E. J. Perreault, "The number and choice of muscles impact the results of muscle synergy analyses," *Front. Comput. Neurosci.*, vol. 7, p. 105, Aug. 2013.

[11] D. Rimini, V. Agostini, and M. Knaflitz, "Intra-Subject Consistency during Locomotion: Similarity in Shared and Subject-Specific Muscle Synergies," *Front. Hum. Neurosci.*, 2017.

[12] H. S. Seung and D. D. Lee, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[13] G. Torres-Oviedo and L. H. Ting, "Muscle Synergies Characterizing Human Postural Responses," *J. Neurophysiol.*, vol. 98, no. 4, pp. 2144–2156, 2007.

[14] S. Hagio, M. Fukuda, and M. Kouzaki, "Identification of muscle synergies associated with gait transition in humans," *Front Hum Neurosci*, vol. 9, no. February, pp. 1–12, 2015.

[15] D. J. Clark, L. H. Ting, F. E. Zajac, R. R. Neptune, and S. A. Kautz, "Merging of healthy motor modules predicts reduced locomotor performance and muscle coordination complexity post-stroke.," *J Neurophysiol*, vol. 103, no. 2, pp. 844–857, 2010.

[16] F. Di Nardo *et al.*, "A new parameter for quantifying the variability of surface electromyographic signals during gait: The occurrence frequency," *J. Electromyogr. Kinesiol.*, vol. 36, pp. 25–33, 2017.

[17] V. Agostini, A. Nascimbeni, A. Gaffuri, P. Imazio, M. G. Benedetti, and M. Knaflitz, "Normative EMG activation patterns of school-age children during gait," *Gait Posture*, vol. 32, no. 3, pp. 285–289, 2010.

[18] F. Di Nardo, E. Maranesi, A. Mengarelli, G. Ghetti, L. Burattini, and S. Fioretti, "Assessment of the variability of vastii myoelectric activity in young healthy females during walking: A statistical gait analysis," *J. Electromyogr. Kinesiol.*, vol. 25, no. 5, pp. 800–807, 2015.

[19] A. Strazza *et al.*, "Surface-EMG analysis for the quantification of thigh muscle dynamic co-contractions during normal gait," *Gait Posture*, vol. 51, pp. 228–233, 2017.

[20] S. Rosati, V. Agostini, M. Knaflitz, and G. Balestra, "Muscle activation patterns during gait: A hierarchical clustering analysis," *Biomed. Signal Process. Control*, vol. 31, pp. 463–469, 2017.

[21] S. Rosati, C. Castagneri, V. Agostini, M. Knaflitz, and G. Balestra, "Muscle contractions in cyclic movements: Optimization of CIMAP algorithm," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 58–61, 2017.

[22] C. Castagneri, V. Agostini, G. Balestra, M. Knaflitz, M. Carlone, and G. Massazza, "EMG Asymmetry Index in Cyclic Movements," in *2018 IEEE Life Sciences Conference (LSC)*, 2018, pp. 223–226.

[23] C. Castagneri, V. Agostini, S. Rosati, G. Balestra, and M. Knaflitz, "Asymmetry Index in Muscle Activations," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 772–779, 2019.

[24] C. Castagneri, V. Agostini, S. Rosati, G. Balestra, and M. Knaflitz, "Longitudinal assessment of muscle function after Total Hip Arthroplasty : Use of clustering to extract principal activations from EMG signals," in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2018, pp. 1–5.

[25] V. Agostini, G. Balestra, and M. Knaflitz, "Segmentation and classification of gait cycles," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 5, pp. 946–952, 2014.

[26] L. Gizzi, S. Muceli, F. Petzke, and D. Falla, "Experimental Muscle Pain Impairs the Synergistic Modular Control of Neck Muscles.," *PLoS One*, vol. 10, no. 9, p. e0137844, 2015.

[27] P. Bonato, T. D'Alessio, and M. Knaflitz, "A statistical method for the measurement of muscle activation intervals from surface myoelectric signal during gait," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 3, pp. 287–299, Mar. 1998.

[28] D. Rimini, V. Agostini, and M. Knaflitz, "Evaluation of muscle synergies stability in human locomotion: A comparison between normal and fast walking speed," *I2MTC 2017 - 2017 IEEE Int. Instrum. Meas. Technol. Conf. Proc.*, no. June 2019, 2017.

[29] K. E. Zelik, V. La Scaleia, Y. P. Ivanenko, and F. Lacquaniti, "Can modular strategies simplify neural control of multidirectional human locomotion?," *J. Neurophysiol.*, vol. 111, no. 8, pp. 1686–1702, Apr. 2014.

[30] M. H. Soomro, S. Conforto, G. Giunta, S. Ranaldi, and C. De Marchis, "Comparison of initialization techniques for the accurate extraction of muscle synergies from myoelectric signals via nonnegative matrix factorization," *Appl. Bionics Biomech.*, vol. 2018, 2018.

[31] A. d'Avella and E. Bizzi, "Shared and specific muscle synergies in natural motor behaviors.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 8, pp. 3076–81, Feb. 2005.

[32] T. Van Criekinge *et al.*, "Lower limb muscle synergies during walking after stroke: a systematic review," *Disabil. Rehabil.*, vol. 0, no. 0, pp. 1–10, 2019.

[33] E. Bizzi and V. C. K. Cheung, "The neural origin of muscle synergies," *Front Comput Neurosci*, vol. 7, no. April, p. 51, 2013.

[34] P. Kieliba, P. Tropea, E. Pirondini, M. Coscia, S. Micera, and F. Artoni, "How are Muscle Synergies Affected by Electromyography Pre-Processing?," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 882–893, 2018.

[35] B. R. Shuman, M. H. Schwartz, and K. M. Steele, "Electromyography Data Processing Impacts Muscle Synergies during Gait for Unimpaired Children and Children with Cerebral Palsy," *Front Comput Neurosci*, vol. 11, no. June, pp. 1–9, 2017.

[36] D. Rimini, V. Agostini, S. Rosati, C. Castagneri, G. Balestra, and M. Knaflitz, "Influence of pre-processing in the extraction of muscle synergies during human locomotion," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 2502–2505, 2017.

[37] S. Ranaldi, C. De Marchis, M. Rinaldi, and S. Conforto, "The effect of Non-Negative Matrix Factorization initialization on the accurate identification of muscle synergies with correlated activation signals,"

MeMeA 2018 - 2018 IEEE Int. Symp. Med. Meas. Appl. Proc., pp. 1–5, 2018.

[38] M. Rinaldi, C. D'Anna, M. Schmid, and S. Conforto, "Assessing the influence of SNR and pre-processing filter bandwidth on the extraction of different muscle co-activation indexes from surface EMG data," *J. Electromyogr. Kinesiol.*, vol. 43, pp. 184–192, 2018.

[39] A. S. P. Sousa and J. M. R. S. Tavares, "Surface electromyographic amplitude normalization methods: A review," *Electromyogr. New Dev. Proced. Appl.*, pp. 85–102, 2012.

[40] V. Agostini *et al.*, "Instrumented Gait Analysis for an Objective Pre-/Postassessment of Tap Test in Normal Pressure Hydrocephalus," *Arch. Phys. Med. Rehabil.*, vol. 96, no. 7, pp. 1235–1241, 2015.

[41] M. G. Benedetti, V. Agostini, M. Knaflitz, V. Gasparroni, M. Boschi, and R. Piperno, "Self-reported gait unsteadiness in mildly impaired neurological patients: an objective assessment through statistical gait analysis.," *J. Neuroeng. Rehabil.*, vol. 9, p. 64, 2012.

[42] V. Agostini, A. Nascimbeni, A. Gaffuri, and M. Knaflitz, "Multiple gait patterns within the same Winters class in children with hemiplegic cerebral palsy," *Clin. Biomech.*, vol. 30, no. 9, pp. 908–914, 2015.

[43] K. L. Rodriguez, R. T. Roemmich, B. Cam, B. J. Fregly, and C. J. Hass, "Persons with Parkinson's disease exhibit decreased neuromuscular complexity during gait," *Clin. Neurophysiol.*, vol. 124, no. 7, pp. 1390–1397, 2013.

[44] J. L. Allen, J. L. McKay, A. Sawers, M. E. Hackney, and L. H. Ting, "Increased neuromuscular consistency in gait and balance after partnered, dance-based rehabilitation in Parkinson's disease," *J. Neurophysiol.*, vol. 118, no. 1, pp. 363–373, 2017.