POLITECNICO DI TORINO
Repository ISTITUZIONALE

The Effect of Number of Gait Cycles on Principal Activation Extraction

Original
The Effect of Number of Gait Cycles on Principal Activation Extraction / Dotti, G., Ghislieri, M., Rosati, S., Agostini, V., Knaflitz, M., Balestra, G.. - ELETTRONICO. - (2021), pp. 985-988. (2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) Mexico 1-5 Nov. 2021) [10.1109/EMBC46164.2021.9629818].

Availability:
This version is available at: 11583/2944632 since: 2021-12-13T09:42:58Z

Publisher:
IEEE

Published
DOI:10.1109/EMBC46164.2021.9629818

Terms of use:

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

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

09 November 2026

The Effect of Number of Gait Cycles on Principal Activation Extraction

G. Dotti, M. Ghislieri, *Student Member*, *IEEE*, S. Rosati, *Member*, *IEEE*, V. Agostini, *Member*, *IEEE*, M. Knaflitz, *Member*, *IEEE*, and G. Balestra, *Member, IEEE*

***Abstract*— To cope with the high intra-subject variability of muscle activation intervals, a large amount of gait cycles is necessary to clearly document physiological or pathological muscle activity patterns during human locomotion. The Clustering for Identification of Muscle Activation Pattern (CIMAP) algorithm has been proposed to help clinicians obtaining a synthetic and clear description of normal and pathological muscle functions in human walking. Moreover, this algorithm allows the extraction of Principal Activations (PAs), defined as those muscle activations that are strictly necessary to perform a specific task and occur in every gait cycle. This contribution aims at assessing the impact of the number of gait cycles on the extraction of the PAs. Results demonstrated no statistically significant differences between PAs extracted considering different numbers of gait cycles, revealing, on average, similarity values higher than 0.88.**

***Clinical Relevance*—This contribution demonstrates the potential applicability of the CIMAP algorithm to the analysis of subjects affected by neurological disorders, for whom the assessment of motor functions may be of the uttermost importance and only a reduced number of gait cycles can be acquired.**

I. Introduction

The analysis of surface electromyographic (sEMG) signals is commonly used to quantitatively assess normal and pathological muscle functions in human walking. Moreover, the assessment of sEMG signals can be a valuable tool in the evaluation of locomotion pathologies and rehabilitation protocols [1]. However, the great stride-to-stride variability of sEMG signals collected during gait [2], [3], even in healthy subjects, may strongly reduce the interpretability and reliability of the results. More specifically, previous studies reported that a subject's walk can be characterized by 4-5 different muscle activation modalities (i.e., number of muscle activation intervals occurring within the same gait cycle) [2], [4] and different patterns within the same modality. Hence, a specific muscle may be activated with a variable sequence of patterns during the walking task. To cope with the high intra-cycle variability of the sEMG signals and to increase the interpretability and reliability of the results, the CIMAP (Clustering for Identification of Muscle Activation Pattern) algorithm was recently developed [5], [6] and validated on several healthy and pathological populations [7]–[9]. The CIMAP algorithm is based on a hierarchical clustering method that allows the extraction of the different patterns. Its application permits the characterization of cyclical movements and the extraction of the Principal Activations (PAs). From a biomechanical point of view, PAs can be defined as those muscle activations that are strictly necessary to perform a specific task [6]. This algorithm allows clustering together the gait cycles showing similar sEMG activation intervals. Each cluster is described by an element (called cluster's prototype) defined as the median of all the muscle activation intervals belonging to the same cluster. Then, PAs are computed as the intersection of all the representative clusters' prototypes. In the optimized version of the CIMAP algorithm, to select the representative clusters, a cutoff threshold (Th) on the number of gait cycles per modality is applied. Clusters with a number of gait cycles per modality higher than this threshold are considered as representative and, hence, used for PA extraction, while clusters with a number of gait cycles per modality lower than this threshold are considered as non-representative [5], [7] and do not contribute to PA extraction. Therefore, the definition of this cutoff threshold is essential. On the one hand, the value of this cutoff threshold may affect the PA extraction. On the other hand, a high value of Th may limit the applicability of this approach to datasets containing a reduced number of gait cycles (e.g., 30-s lasting protocols [10]).

The aim of this contribution is to assess the robustness of the cutoff threshold and to assess the effect of the number of gait cycles on the principal activation extraction, enhancing the applicability of the CIMAP algorithm to shorter acquisition protocols and different cohorts (i.e., patients affected by musculoskeletal or neurological disorders).

II. Materials and Methods

A. Sample Population and Data Acquisitions

Gait data from 20 healthy subjects (11 males and 9 females, age: 65.4 ± 5.1 years, height: 1.69 ± 0.09 m, weight: 69.0 ± 12.2 kg) were retrospectively analyzed from our gait data warehouse (BIOLAB, Politecnico di Torino, Turin, Italy). None of the subjects had neurological or orthopedic pathologies that could affect gait performance. Gait data were recorded using a multichannel acquisition system for gait analysis (STEP32, Medical Technology, Italy) with a sampling frequency of 2 kHz. More specifically, sEMG, foot-

Gregorio Dotti, Marco Ghislieri, Samanta Rosati, Valentina Agostini, Marco Knaflitz, and Gabriella Balestra are with the Department of Electronics and Telecommunications, Politecnico di Torino, 10129, Turin, and also with PoliToBIOMed Lab, Politecnico di Torino, 10129, Turin, Italy (e-mail: gregorio.dotti@studenti.polito.it, marco.ghislieri@polito.it, samanta.rosati@polito.it, valentina.agostini@polito.it, marco.knaflitz@polito.it, and gabriella.balestra@polito.it)

switch, and joint kinematic signals were simultaneously acquired during gait. SEMG signals were recorded through active probes placed bilaterally over the following 5 lower limb muscles: Rectus Femoris (RF), Lateral Hamstring (LH), Gluteus Medius (GMD), Tibialis Anterior (TA), and Lateral Gastrocnemius (LGS). Foot-switches were positioned under the foot sole of both lower limbs to time-segment gait cycles. Finally, electrogoniometers were placed on the lateral aspect of both the knee joints to assess joint kinematics in the sagittal plane.

All the enrolled subjects underwent the same experimental protocol consisting of an instrumented 3-minutes walk on a 10-m straight walkway, at a self-selected speed. All participants signed written informed consent for the experimental procedure and all the experiments were performed following the Declaration of Helsinki.

B. *Data Pre-Processing*

Before the application of the CIMAP algorithm, sEMG signals were pre-processed to extract muscle activation intervals and to time-segment all gait cycles.

To extract the muscle activation intervals, sEMG signals were first pass-band filtered through a 5th order Butterworth digital filter with a lower cut-off frequency of 40 Hz and an upper cut-off frequency of 300 Hz to remove motion and high-frequency artifacts [11], respectively. Then, muscle activation intervals were extracted from the filtered sEMG signal using a muscle activity detector based on Long Short-Term Memory (LSTM) recurrent neural networks [12]–[14]. The muscle activation intervals computed through the muscle activity detector were defined as binary masks that were set equal to 1 when the muscle was identified as active and 0 otherwise. To avoid erroneous transitions due to the stochastic nature of the sEMG signals, a post-processing step was applied to the detector's output to remove activation intervals shorter than 30 ms [15].

For each acquired muscle, muscle activation intervals were time-segmented into gait cycles using the initial contact timings extracted from the foot-switch signals [16]. Finally, to remove differences in gait cycle duration, all the segmented gait cycles were time-normalized to 1000 samples using the Linear Length Normalization (LLN) approach [17]–[19].

C. *Extraction of Principal Activations (PAs) through CIMAP Algorithm*

The optimized version of the CIMAP algorithm [5] was applied to the normalized muscle activation intervals to extract PAs, separately for each muscle. The computation of the PAs using CIMAP algorithm mainly consists of 5 phases:

i. Bottom-up hierarchical clustering is applied to the time-normalized sEMG activation intervals, separately for each muscle. The CIMAP algorithm iteratively merges the two "closest" clusters considering the Chebyshev and Manhattan distances, separately. The complete linkage method is used to select whether two clusters have to be merged;

ii. The result of the bottom-up hierarchical clustering is represented by two different dendrograms. The final number of clusters (cutoff rule) is then selected, separately for each dendrogram, to obtain clusters characterized by small intra-cluster variability (ICV). The intra-cluster variability is defined as the Euclidean

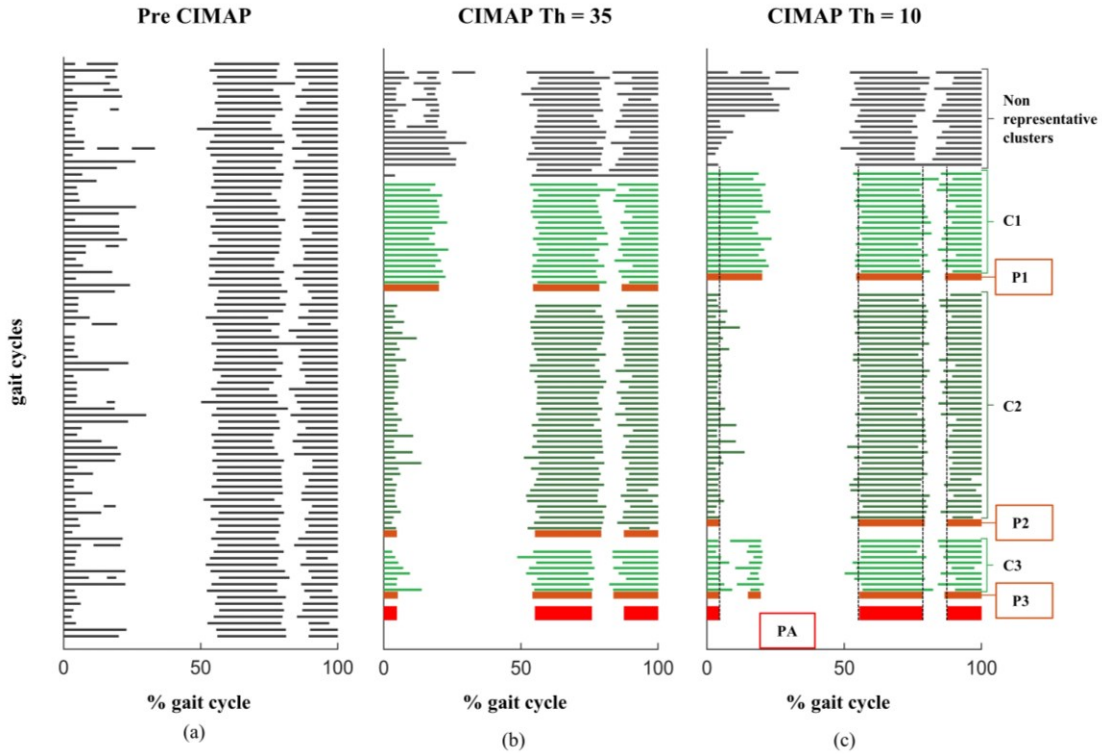

Figure 1: Example of application of the CIMAP algorithm to the Tibialis Anterior (TA) muscle of a representative subject of the sample population. Panel (a) represents the time-normalized sEMG activation intervals before CIMAP application (gray lines). Panels (b) and (c), instead, represent sEMG activation intervals grouped into clusters (green lines), the clusters' prototypes (orange lines), and the PAs (red lines) extracted considering two different values of the CIMAP threshold, respectively. Gray lines on top of panels (b) and (c) represents the non-representative clusters.

distance between each element of the cluster and the corresponding cluster's prototype;

iii. The cutting point is obtained using the difference between the variability of the two clusters that are merged and the resulting cluster. Three different cutting points are computed for each dendrogram and the best one is automatically chosen as described in [5];

iv. Non-representative clusters (i.e., clusters with a number of gait cycles per modality lower than a cutoff threshold - Th) are discarded [5];

v. PAs are then extracted from all the representative clusters as the intersection of the clusters' prototypes [5]. PAs are defined as binary masks that are set equal to 1 in correspondence of a principal activation and to 0 otherwise.

Figure 1 shows an example of the application of the CIMAP algorithm on the muscle activation intervals extracted from the Tibialis Anterior (TA) muscle of a representative subject of the sample population. Figure 1(a) represents the time-normalized sEMG activation intervals (gray lines) before the application of the CIMAP algorithm. Figure 1(b) and Figure 1(c), instead, represent the time-normalized sEMG activation intervals (green lines) grouped in clusters with the indication of the clusters' prototypes (orange lines) and PAs (red lines) extracted considering two different values of the CIMAP threshold (Th), respectively.

D. Robustness of the CIMAP Threshold

To assess if the extraction of PAs is robust with respect to the CIMAP threshold Th, the following procedure was followed, separately for each muscle. We compared the PAs extracted from the time-normalized muscle activation intervals of the whole dataset, considering 6 different values of the CIMAP threshold: Th = 35, 30, 25, 20, 15, and 10 gait cycles per modality. These values of the CIMAP threshold Th have been chosen to obtain a sufficient number of gait cycles per modality for the subsequent extraction of PAs.

The PAs extracted using Th_{35}, Th_{30}, Th_{25}, Th_{20}, Th_{15}, and Th_{10} were quantitatively compared by computing the similarity index (D) between each pair of PAs as defined in (1):

$$D = 1 - \frac{\sum_{i=1}^{n} \left| PA_{A,i} - PA_{B,i} \right|}{n} \qquad (1)$$

where n represents the number of samples within a gait cycle (i.e., 1000 time-samples) and $PA_{A,i}$ and $PA_{B,i}$ represent the principal activations extracted from the i-th subject considering two different CIMAP thresholds (A and B), respectively. The similarity index D ranges from 0 (i.e., no similarity) to 1 (i.e., complete similarity).

E. Effect of the Number of Gait Cycles on Principal Activation Extraction

To assess the effect of the number of gait cycles on principal activation extraction, the CIMAP algorithm was applied to three different datasets, each of them characterized by a different number of gait cycles:

i. *Dataset A*: it contains all gait cycles recorded during the experimental sessions. On average, 153 ± 18 gait cycles were acquired from each subject of the sample population;

ii. *Dataset B*: it contains, on average, 72 ± 9 gait cycles for each subject, corresponding to the gait cycles acquired during the first 10 10-m straight walkways;

iii. *Dataset C*: it contains, on average, 35 ± 2 gait cycles for each subject, corresponding to the gait cycles acquired during the first 5 10-m straight walkways.

These datasets, built to simulate increasingly shorter experimental protocols, were then used as inputs of the CIMAP algorithm to extract PAs. The effect of the number of gait cycles on principal activation extraction was quantitatively assessed computing the similarity index described in the previous paragraph (*D. Robustness of the CIMAP Threshold*) between PAs computed considering the three tested datasets.

F. Statistical Analysis

First, the Lilliefors test was used to assess the normality of data distributions. Second, depending on the results of the Lilliefors test, one-way repeated measures ANOVA (normal distributions) or Kruskal-Wallis test (non-normal distributions) was applied to test statistically significant differences between PAs. Finally, *post-hoc* analysis with Tukey's adjustments for multiple comparisons was implemented. The statistical analysis was performed using the Statistical and Machine Learning Toolbox of MATLAB® release R2020b (The MathWorks Inc., Natick, MA, USA), setting the significance level (α) equal to 0.05.

III. Results and Discussion

First, we present the results related to the robustness of the CIMAP threshold (Th). Then, we present the results related to the analysis of the impact of the number of gait cycles on the principal activation extraction.

To avoid the detection of erroneous muscle activation intervals due to low-quality sEMG signals, sEMG data revealing Signal-to-Noise Ratio values lower than 6 dB were discarded. According to this cut-off threshold, 3 out of 20 subjects of the sample population were excluded and, hence, the following results were extracted from the remaining 17 healthy subjects.

TABLE I. SIMILARITY VALUES (D) BETWEEN PAS EXTRACTED CONSIDERING DIFFERENT NUMBER OF GAIT CYCLES.

	Muscles					
	TA	***LGS***	***RF***	***LH***	***GMD***	***All Muscles***
Dataset A vs. Dataset B	0.93 ± 0.09	0.93 ± 0.06	0.93 ± 0.06	0.91 ± 0.07	0.95 ± 0.04	0.93 ± 0.03
Dataset A vs. Dataset C	0.90 ± 0.17	0.92 ± 0.17	0.90 ± 0.17	0.89 ± 0.09	0.91 ± 0.17	0.90 ± 0.08
Dataset B vs. Dataset C	0.90 ± 0.16	0.90 ± 0.17	0.90 ± 0.17	0.88 ± 0.09	0.91 ± 0.17	0.90 ± 0.07

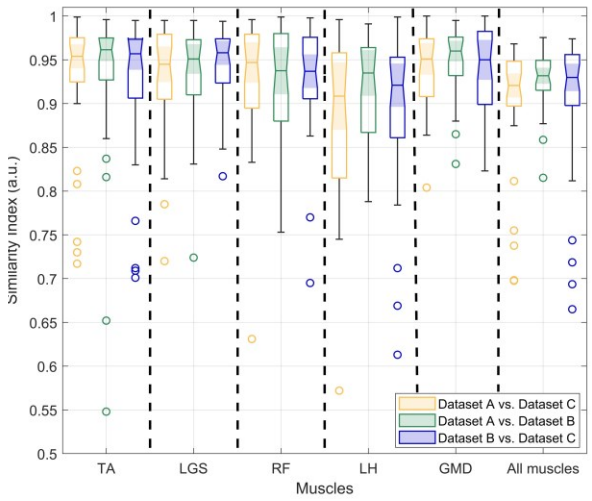

Figure 2. Boxplots of the similarity indexes D relative to each of the 5 acquired muscles and to the grand average over all the muscles, separately for each of the three tested datasets. Outliers are represented through circles.

i. Robustness of the CIMAP Threshold

Results revealed high values of the similarity index ($D >$ 0.94) between all the tested CIMAP thresholds (Th_{35}, Th_{30}, Th_{25}, Th_{20}, Th_{15}, and Th_{10}). Kruskal-Wallis's test followed by *post-hoc* analysis with Tukey's adjustments for multiple comparisons revealed statistically significant differences ($p <$ 0.01) between PAs extracted considering Th_{30}, Th_{15}, and Th_{10}. However, considering the high similarity values between the PAs extracted considering the 6 different thresholds, the CIMAP threshold Th was set equal to 10 gait cycles to allow the application of the CIMAP algorithm also to shorter acquisition protocols.

ii. Effect of the Number of Gait Cycles on Principal Activation Extraction

Table I shows the values of the distance metric (D), averaged on the sample population, computed between each pair of PAs extracted considering the three different datasets. Kruskal-Wallis's test revealed no statistically significant differences between the three tested datasets ($p > 0.62$), suggesting a low effect of the number of gait cycles on the extraction of PAs. Figure 2 shows the boxplots of the similarity indexes relative to each of the 5 acquired muscles and the boxplot of the grand average over all the muscles. Considering the high similarity obtained between the PAs extracted from the 3 tested datasets, the CIMAP algorithm should be effectively applied to datasets containing a reduced number of gait cycles (e.g., 30-s lasting protocols).

IV. Conclusion

Results presented in this contribution demonstrated that PAs extracted from healthy subjects during a walking task at comfortable walking speed are not affected by the number of gait cycles considered as input of the CIMAP algorithm. Nevertheless, further studies are necessary to test the effect of the number of gait cycles also on sEMG signals acquired from pathological sample populations, such as patients affected by musculoskeletal or neurological disorders.

References

[1] V. Agostini, M. Ghislieri, S. Rosati, G. Balestra, and M. Knaflitz, "Surface Electromyography Applied to Gait Analysis: How to Improve Its Impact in Clinics?," *Front. Neurol.*, vol. 11, pp. 1–13, 2020, doi: 10.3389/fneur.2020.00994.

[2] 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, doi: https://doi.org/10.1016/j.gaitpost.2010.06.024.

[3] F. Di Nardo and S. Fioretti, "Statistical analysis of surface electromyographic signal for the assessment of rectus femoris modalities of activation during gait," *J. Electromyogr. Kinesiol.*, vol. 23, no. 1, pp. 56–61, Feb. 2013, doi: 10.1016/J.JELEKIN.2012.06.011.

[4] F. Di Nardo *et al.*, "Normative EMG patterns of ankle muscle co-contractions in school-age children during gait," *Gait Posture*, vol. 46, pp. 161–166, 2016, doi: https://doi.org/10.1016/j.gaitpost.2016.03.002.

[5] 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, doi: 10.1109/EMBC.2017.8036762.

[6] 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, doi: 10.1016/j.bspc.2016.09.017.

[7] V. Agostini, S. Rosati, C. Castagneri, G. Balestra, and M. Knaflitz, "Clustering analysis of EMG cyclic patterns: A validation study across multiple locomotion pathologies," in *2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, May 2017, pp. 1–5, doi: 10.1109/I2MTC.2017.7969746.

[8] 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, doi: 10.1109/MeMeA.2018.8438802.

[9] 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, doi: 10.1109/LSC.2018.8572041.

[10] M. W. Flood, B. P. F. O'Callaghan, and M. M. Lowery, "Gait Event Detection from Accelerometry Using the Teager-Kaiser Energy Operator," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 3, pp. 658–666, Mar. 2020, doi: 10.1109/TBME.2019.2919394.

[11] C. J. De Luca, L. Donald Gilmore, M. Kuznetsov, and S. H. Roy, "Filtering the surface EMG signal: Movement artifact and baseline noise contamination," *J. Biomech.*, vol. 43, no. 8, pp. 1573–1579, 2010, doi: 10.1016/j.jbiomech.2010.01.027.

[12] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[13] I. . Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *J. Microbiol. Methods*, vol. 43, no. 1, pp. 3–31, Dec. 2000, doi: 10.1016/S0167-7012(00)00201-3.

[14] A. Graves, "Generating Sequences With Recurrent Neural Networks," pp. 1–43, 2013, [Online]. Available: http://arxiv.org/abs/1308.0850.

[15] 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, doi: 10.1109/10.661154.

[16] V. Agostini *et al.*, "Segmentation and classification of gait cycles," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 5, pp. 946–952, 2014, doi: 10.1109/TNSRE.2013.2291907.

[17] N. E. Helwig, S. Hong, E. T. Hsiao-Wecksler, and J. D. Polk, "Methods to temporally align gait cycle data," *J. Biomech.*, vol. 44, no. 3, pp. 561–566, 2011, doi: 10.1016/j.jbiomech.2010.09.015.

[18] D. A. Winter, *Biomechanics and motor control of human movement*. Wiley, 2009.

[19] S. R. Simon, "Gait Analysis, Normal and Pathological Function.," *J. Bone Jt. Surg.*, vol. 75, no. 3, pp. 476–477, Mar. 1993, doi: 10.2106/00004623-199303000-00027.