Influence of EMG-signal processing and experimental set-up on prediction of gait events by neural network

Influence of EMG-signal processing and experimental set-up on prediction of gait events by neural network

Francesco DI NARDO[1], Christian MORBIDONI[1],
Alessandro CUCCHIARELLI[1], Sandro FIORETTI[1]

[1] Department of Information Engineering, Università Politecnica delle Marche,
via Brecce Bianche, 60131 Ancona, Italy.

Corresponding author:
Francesco Di Nardo, Ph.D.
Department of Information Engineering
Università Politecnica delle Marche
Via Brecce Bianche, 60131 Ancona, Italy
Fax: (39)0712204224; e-mail: f.dinardo@staff.univpm.it

Abstract

Machine-learning approaches are satisfactorily implemented for classifying and assessing gait events from only surface electromyographic (sEMG) signals during walking. However, it is acknowledged that the choice of sEMG-processing type may affect the reliability of methodologies based on it. Analogously, the number of sEMG signals involved in machine-learning procedure could influence the classification process. Aim of this study is to quantify the impact of different EMG-signal-processing specifications and/or different complexity of the experimental sEMG-protocol (different number of sEMG-sensors) on the performance of a neural-network-based approach for binary classifying gait phases and predicting gait-event timing. To this purpose, sEMG signals are collected from eight leg-muscles in about 10.000 strides from 23 healthy adults during walking and then fed to a multi-layer perceptron model. Four different signal-processing approaches are tested and five experimental set-ups (from four to one sEMG sensors per leg) are compared. Results indicate that both the choice of sEMG processing and the reduction of sEMG-protocol complexity actually affect classification/prediction performances. Moreover, the study succeeds in the double goal of identifying the linear envelope as the sEMG-processing type which reaches the best neural-network performance (classification accuracy of 93.4±2.3%; mean absolute error 21.6±7.0 and 38.1±15.2 ms for heel-strike/toe-off prediction, respectively) and providing a quantification of the progressive deterioration of classification/prediction performances with the reduction of the number of sensors used (from 93.4±2.3% to 79.9±6.1% for classification accuracy). These findings could be very useful for clinics to the aim of choosing the most suitable approach balancing technical performances, patient comfort, and clinical needs.

1. Introduction

Each single gait cycle of human walking is composed of two main phases: the stance phase, from the beginning to around 60% of gait cycle; the swing phase, from 60% to the end of gait cycle [1]. The stance phase denotes the whole time interval when the reference foot is touching the ground. The swing phase quantifies the period when the foot is no longer on the ground and swings in the air for leg advancement. Crucial for quantification of gait phases duration are the transition events between the two phases: toe-off (TO, from stance to swing) and heel-strike (HS, from swing to stance). The assessment of these temporal parameters is one of the typical tasks of gait analysis [2,3].

In the recent years, artificial-intelligence techniques have been proposed for the classification of stance vs. swing and for the assessment of temporal gait events [4,5]. Particularly valuable are those methodologies where machine and deep learning are implemented with the aim of limiting the number of sensors involved in the experimental set-up, such as electromyography-based approaches [6-12]. These studies are designed to classify gait phases and predict gait events from only surface electromyographic signals (sEMG), avoiding the requirement of directly measuring temporal data by means of additional systems or sensors (foot-switch sensors, IMUs, pressure mats, stereo-photogrammetry). This would allow to reduce burden for patient, simplify clinical protocols, and make test faster, specifically in the evaluation of neuromuscular diseases or for walking-aid devices where the acquisition of myoelectric signals is largely advised [13,14]. The advantage would be even greater if it could be possible not only limiting the number of sensors for temporal-data measurement, but also decreasing the number of sEMG probes themselves. Obviously, reducing the number of sEMG sensors means having fewer signals to be processed by the neural network. This is expected to lead to a deterioration of classification performances. To our knowledge, a reliable analysis of the effect on classification/prediction performance of the reduction of sEMG signals involved in feeding the neural network is not yet available in literature.

Furthermore, the problem of gait-phase classification with neural-network-based interpretation of only electromyographic signals has been typically faced extracting explicit features from sEMG signal and then using them as input to the machine learning stages [6-10]. The present group of researchers recently experimented a different strategy [11,12], consisting in the application of neural networks to learn hidden features from a processed sEMG signal. This strategy seems to improve the classification performances [11], but at the same time it introduces the need of identifying the most suitable sEMG-processing type. Recent studies, indeed, indicate that the choice of processing type and processing-parameter value could be very subjective [15], could influence the reliability of methodologies implemented to assess muscle activity [16], and could also affect the estimation of gait events (HS and TO) [17]. Thus, the choice of the sEMG processing is still an open issue.

The aim of the present study is to quantify the impact of different complexity of the experimental sEMG-protocol (i. e. different number of sEMG sensors) and/or different EMG-signal-processing specifications on the performance of a neural-network-based approach for the binary classification of gait phases and prediction of gait-event (HS and TO) timing. This aim is pursued, attempting to provide the following main contributions:

1) identifying which one of the following widely-used approaches to process EMG signals allows to achieve the best classification/prediction performances: a) band-pass filtered signal; b) full-wave rectified signal; c) linear envelope of the signal; and d) root mean square signal. Details of these sEMG-signal processing are reported in Section 3.3;

2) testing the sensitivity of the performances to different values of envelope cut-off-frequency. Values of 5, 10, 15, and 20 Hz were adopted, considering that envelope cut-off frequency typically ranges from 3 Hz to 25 Hz [15].
3) quantifying the conceivable decrease of classification/prediction performance with the reduction of the number of sEMG signals involved in feeding the neural network. Five experimental set-ups are considered to this purpose, including: 1) sensors positioned on the proximal and distal leg (medial hamstrings, MH, vastus lateralis, VL, tibialis anterior, TA, and gastrocnemius lateralis, GL, full set-up; 2) only sensors positioned on the proximal leg (MH and VL, proximal leg set-up); 3) only sensors positioned on the distal leg (TA and GL, distal leg set-up); 4) only sensors positioned on tibialis-anterior muscle (TA set-up); and 5) only sensors positioned on gastrocnemius-lateralis muscle (GL set-up).

The manuscript is organized as follows: Section 2 provides a brief review of the related works. Section 3 introduces the dataset, illustrates the acquisition and the pre-processing of the signals, describes the procedure of gait-phase classification and gait-event prediction by machine-learning approach, and presents the statistical tests. Section 4 reports the experimental results that are then discussed in Section 5. Both results and discussion sections are split in two sub-sections about signal pre-processing and reduction of experimental set-up, respectively. Eventually, Section 6 ends the study and provides insights for further research developments.

2. Related Works

A relatively small number of works in literature address gait-phase classification from EMG signals only. In [6] a set of time-domain features, namely mean absolute value (MAV), waveform length (WL), zero crossing (ZC), and slope sign changes (SSC) were extracted from EMG signal and hidden Markov models were used to classify stance and swing phases. Evaluation on treadmill walking of a single subject reported a maximum accuracy of 91.1%. Monitored muscles are Vastus Medialis, Semitendinosus, Adductors, and Tensor Fascia Latae. A novel bilateral EMG feature, called weighted signal difference (WSD), was introduced in [9] and used to train a support vector classifier (SVC). Intra-subject evaluation is performed on two subjects walking on a treadmill at different speed, reporting a best accuracy of 96%. Monitored muscles were Soleus, Tibialis Anterior, Gastrocnemius Lateralis, Vastus Lateralis, Rectus Femoris and Gluteus Maximus. In [7] a control system for a foot-knee exoskeleton based on the processing of eight EMG signals is proposed. Four time-domain features (MAV, WL, Variance and SGC) were extracted and Bayesian Information Criteria (BIC) was used to predict 8 distinct gait events. Evaluation on one healthy subject revealed low repeatability of the method, with a 30% drop in accuracy testing on different gait cycles. Monitored muscles were Quadriceps, Hamstring, Gastrocnemius and Tibialis Anterior. In [8] and [10] a set of temporal features, namely root mean square (RMS), standard deviation (SD), MAV, WL, and integrated EMG (IEMG), were fed to a single layer neural networks to detect TO and HS on a population of 8 healthy adults. The study targets inter-subject prediction by testing the network on one unlearned subject (not used in training), however no cross validation is performed and the test is performed on a 5-second walk only. No indication is provided regarding accuracy of prediction and a mean average error of 35 ms and 49 ms is reported for HS and TO prediction respectively. Monitored muscles were Tibialis Anterior (TA) and Medial Gastrocnemius (mGas).

All the works mentioned above were based on explicit features extraction and used different sets of features as input to the machine learning stage. Recently, a different approach was introduced [11,12], where the original sEMG signal is first pre-processed, in order to obtain a smoothed and cleaner signal, and then neural networks were used to learn hidden features, classifying the two main gait phases and successively individuate the TO and HS events as the transitions between different phases. The sEMG signals acquired during level ground walking from eight lower-limb muscles, tibialis anterior (TA), gastrocnemius lateralis (GL), medial hamstrings (MH), and vastus lateralis (VL) of each leg, in more than 10.000 strides from 23 healthy adult subjects were involved [11]. As far as we know, this work is still reporting the best performances in HS and TO prediction (mean absolute error of 21.6 ± 7.0 ms and 38.1 ± 15.2 ms, respectively and F1-score ≈ 99%) among the mentioned sEMG-based studies. These promising results were achieved by feeding the classifier with the envelope of EMG signal, computed as follows [11]: sEMG signal was band-pass filtered (linear-phase FIR filter, cut-off frequency: 20 - 450 Hz), then full-wave rectified, and eventually the envelope was extracted (second-order Butterworth low-pass filter, cut-off frequency: 5 Hz). Such a pre-processing pipeline was designed following the indications provided by previous acknowledged studies [18,19].

3. Materials and Methods

3.1 Participants

Twenty-three able-bodied adults were involved in the experimental procedure. Volunteer data, reported as mean value ±SD, are the following: height = 173 ±10 cm;

mass = 63.3 ± 12.4 kg; age = 23.8 ±1.9 years; and female/male ratio = 12/11. Subjects with articular pain, with disorder of the nervous system, in obese or overweight condition (body mass index > 25), and with history of orthopaedic surgery that may affect walking performances were exempted from the study. The research presented here was undertaken following the ethical principles of Helsinki Declaration and was approved by local ethical committee.

3.2 Signal acquisition

The multichannel recording system Step32 (Medical Technology, Italy, Version PCI-32 ch2.0.1. DV) was employed for signal acquisition (resolution: 12 bits; sampling rate: 2 kHz). Three foot-switches were placed under the heel, the first and the fifth metatarsal heads of both subject's feet, for acquiring foot-floor-contact signal. Four sEMG sensors were applied over vastus lateralis (VL), medial hamstrings (MH), tibialis anterior (TA), and gastrocnemius lateralis (GL) of both legs, complying with recommendations suggested by SENIAM standards [19]. After that, subject walked barefoot approximately 5 minutes on an eight-shaped path at her/his own pace. Experiments were performed in Motion Analysis Laboratory of the Università Politecnica delle Marche, Ancona, Italy. Characteristics of sEMG single-differential probes are: material = Ag/Ag-Cl disks; gain = 1000, filtering = high-pass filter with cut-off frequency of 10 Hz; input impedance > 1.5GΩ; Common-Mode Rejection Ratio > 126 dB; input referred noise ≤ 1 μVrms; and manufacturer = Medical Technology, Italy. sEMG probes with fixed geometry have size of 7 × 27 × 19 mm; electrode diameter of 4 mm; and inter-electrode distance of 8 mm). sEMG probes with variable geometry have a minimum inter-electrode distance of 12 mm. Characteristics of foot-switches are: size = 11 mm × 11 mm × 0.5 mm and activation force = 3 N. Additional information about signal acquisition could be obtained in [20].

3.3 Signal pre-processing

Foot-switch signals were processed for recognizing gait cycles and stance/swing phases [21]. To test the effect of signal filtering on classification performance, sEMG signal were pre-processed with four different approaches.

Band-pass filtered signal (BPFS): sEMG signals were band-pass filtered (linear-phase FIR filter, cut-off frequency: 20 - 450 Hz) for taking out high frequency noise and motion artefacts.

Full-wave Rectified signal (FWRS): sEMG signals were band-pass filtered (linear-phase FIR filter, cut-off frequency: 20 - 450 Hz). Then, a full-wave rectification was achieved, taking the absolute value of the signal.

Linear envelope of the signal (LE): sEMG signals were band-pass filtered (linear-phase FIR filter, cut-off frequency: 20 - 450 Hz) and full-wave rectified. Then, envelope of the signal was extracted (second-order Butterworth low-pass filter). Four different values of cut-off frequency were tested: 5, 10, 15, and 20 Hz. These four different processing of the envelope have been referred to as LE_5, LE_{10}, LE_{15}, and LE_{20}, respectively.

Root mean square signal (RMSS): sEMG signals were band-pass filtered (linear-phase FIR filter, cut-off frequency: 20 - 450 Hz). Then, a sliding window of length *N* scans the signal sample by sample. *RMSV* computed in the first window is the first sample of the Root mean square signal. *RMSV* computed in the second window is the second sample of the Root mean square signal and so on. *RMSV* is computed as in the following formula:

$$RMSV = \sqrt{\frac{1}{N}\sum_{k=1}^{N}|x_k^2|} \quad (1)$$

where N is number of samples, x_k is the k-sample. Two different values of sliding-window duration were tested: 100 samples ($RMSS_{100}$) and 500 samples ($RMSS_{500}$). After every kind of filtering, sEMG signals were min-max normalized within each subject and for each muscle, thus mapping the values in the [0–1] interval. All the pre-processing operations were implemented using Matlab relying on standard functions provided by the Signal Processing Toolbox[1].

3.4 Data preparation

Each sEMG signal was separated into 20-sample windows, matching 10 milliseconds (ms). A chronological sequence of vectors made up of $20 \times n$ samples was composed, where each vector included n synchronized 20-sample windows from sEMG signals of n muscles (n/2 for each leg). In details, the first sample of the first vector of the sequence was the first sample of the sEMG signal from the muscle 1 (TA, right leg), the second sample of the first vector was the first sample of the EMG signal from the muscle 2 (GL, right leg), and so on up to the muscle n. After that, each vector was given a specific label of 0 (or 1), when all basographic-signal samples assume a value of 0 (or 1). Vectors containing transitions between phases were not included in the training set.

The classifier was then trained following the leave-one-out cross validation procedure: 22 out of 23 subjects were involved in training the classifier (Learned subjects, LS); the remaining subject was employed to test the classification output (Unlearned subject, US). LS were further separated into two subsets: training set containing the first 90% of each subject signal (LS-train); testing set including the remaining 10% (LS-test).

[1] https://it.mathworks.com/products/signal.html

In details, the classifier was fed with the vectors extracted from LS-train. The vectors from US and LS-test were employed for testing the classifier performances in unseen subject and in unseen samples of learned subjects, respectively. In this stage, foot-switch signal was the ground truth. This process has been repeated twenty-three times, each time employing a different subject as US.

3.5 The neural network

A Multi Layer Perceptron (MLP) classifier was used in the present study. The MLP architecture is characterized by:

- 3 hidden layers of 512, 256 and 128 neurons, respectively;
- a one-dimensional binary output, provided by applying a 0.5 threshold to a sigmoid activation;
- a rectified linear units (ReLU) between each couple of consecutive hidden layers to supply non-linearity;
- a stochastic gradient descent optimization algorithm with binary cross entropy loss function.

The specific architecture was chosen among others, with different numbers of layers, tested in [11], as it provided the best classification accuracy. In training the network, 10% of the training dataset was used as validation set. At each training epoch, accuracy on the validation set is measured and training was stopped when the validation accuracy did not increase in 10 consecutive epochs. Then, the trained network weights, corresponding to the best validation accuracy, were used to evaluate the model on the LS-test and US sets.

The neural network and the corresponding training and testing code were implemented in Python using the Pytorch deep learning framework[2] and the Scikit-Learn python library[3].

3.6 Gait-event identification

The binary output of the classifier has been chronologically arranged to provide the predicted foot-floor-contact signal, as a vector made up of sequences of 0 (stance phase) alternating with sequences of 1 (swing phase). While signal windows containing transitions were discarded in the training phase, all of them were fed as input to the classifier, in order to predict the foot-floor-contact signal. It is acknowledged that stance and swing are typically lasting around 60% and 40% of gait cycle, during able-bodied walking [1]. Accordingly, the sample sequences shorter than 500 samples (< 25% of gait cycle) were removed to clean out the predicted signal. Afterward, stance-to-swing (toe off, TO) and swing-to-stance (heel strike, HS) transitions have been assessed in the cleaned signal. TO was identified as the sample when the value switched from 0 to 1. Similarly, HS was identified as the sample when the value switched from 1 to 0. Prediction performances were quantified in terms of precision, recall, and F1-score. Precision is computed as:

$$Precision = \frac{TP}{TP + FP} \qquad (2)$$

where *TP* is true positive and *FP* is false positive. Recall is computed as:

---

[2] https://pytorch.org/

[3] https://scikit-learn.org/stable/

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

where FN is false negative. F1-score is computed as:

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \qquad (4)$$

Predicted HS or TO at time t_p were acknowledged as true positives (TP) if an event of the same type occurs in ground-truth signal at time t_g such that $|t_g - t_p| <$ T. T is a time tolerance, set to 600 samples. Otherwise, the predicted event was acknowledged as a false positive (FP). For all TP, mean absolute error (MAE) has been computed, as the mean time distance between the predicted event and the corresponding event in ground-truth signal.

3.7 Statistics

Shapiro-Wilk test was used to evaluate the hypothesis that each data vector had a normal distribution. Comparison between two normally distributed samples was performed with two-tailed, non-paired Student's t-test. Two-sample Kolmogorov-Smirnov test was used to compare not normally distributed samples. Statistical significance was set at 5%.

4. Results

4.1 Signal pre-processing

Average classification accuracies as result of different pre-processing of the signal are reported in Table 1, for Learned-test set (LS-test) and Unlearned set (US). Linear envelope (*LE*) of the signal is evaluated considering four different values of cut-off frequency: 5, 10, 15, and 20 Hz.

Table 1. Mean classification accuracy as result of different pre-processing of the signal.

	Mean classification accuracy (%)	
	LS-test	**US**
LE_5	94.8 ± 0.2	93.4 ± 2.3
LE_{10}	93.8 ± 0.3	93.1 ± 2.4
LE_{15}	93.2 ± 0.3	91.4 ± 2.4
LE_{20}	92.4 ± 0.4	90.3 ±3.3
$RMSS_{100}$	92.3 ± 0.5	90.1 ± 2.9
$RMSS_{500}$	93.0 ± 0.4	91.0 ± 3.7
FWRS	88.8 ± 0.2	88.0 ± 2.9
BPFS	86.5 ± 0.6	84.0 ± 3.7

Classification results highlight that accuracy is decreasing with increasing cut-off frequency in both LS-test (from 94.8% to 92.4%) and US (from 93.4% to 90.3%). In a similar way, SD is increasing with increasing cut-off frequency (from 0.2 to 0.4 in LS-test; from 2.3 to 3.3 in US). Comparison between Root mean square signals (*RMSS*) computed with two different values of sliding-window duration shows slightly better accuracy for $RMSS_{500}$ for both LS-test and US. All *LE* and *RMSS* approaches report a mean classification accuracy > 92% in LS-test and > 90% in US. Otherwise, *FWRS* and

BPFS approaches remain definitely < 90%, in particular for US. Overall, best mean accuracy (and SD) is provided by LE_5 in both LS-test and US.

Table 2. MAE (mean absolute error), precision, recall, and F1-score as result of different pre-processing of the signal for Heel Strike (HS) and Toe Off (TO) prediction in US.

Mean prediction performances				
Heel Strike (HS)	**MAE (ms)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
LE_5	21.6 ± 7.0	99.7 ± 0.6	98.5 ± 3.0	99.0 ± 1.7
LE_{10}	26.7 ± 9.8	99.6 ± 0.7	98.8 ± 1.6	99.2 ± 1.1
LE_{15}	27.4 ± 11.7	99.5 ± 0.6	98.7 ± 1.3	99.1 ± 0.9
LE_{20}	35.1 ± 26.5	98.9 ± 2.5	98.1 ± 3.3	98.5 ± 2.8
$RMSS_{100}$	28.1 ± 9.6	99.2 ± 1.2	98.0 ± 3.0	98.6 ± 1.9
$RMSS_{500}$	33.9 ± 14.3	98.7 ± 2.2	98.1 ± 2.6	98.4 ± 2.3
FWRS	47.3 ± 24.9	99.2 ± 1.0	98.4 ± 1.8	98.9 ± 1.4
BPFS	77.4 ± 40.4	95.6 ± 5.9	90.4 ± 13.0	92.6 ± 9.7
Toe Off (TO)	**MAE (ms)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
LE_5	38.1 ± 15.2	99.1 ± 1.5	97.9 ± 3.6	98.4 ± 2.4
LE_{10}	46.0 ± 22.6	98.7 ± 2.3	97.9 ± 2.9	98.3 ± 2.5
LE_{15}	47.9 ±19.3	98.4 ± 2.3	97.6 ± 2.7	98.0± 2.4
LE_{20}	58.2 ± 26.4	98.5 ± 2.1	97.6 ± 2.8	98.0 ± 2.4
$RMSS_{100}$	58.3 ± 22.3	98.6 ± 1.9	97.4 ± 3.6	97.9 ± 2.7
$RMSS_{500}$	54.1 ± 29.5	97.8 ± 3.2	97.1 ± 3.6	97.5 ± 3.4
FWRS	58.8 ± 29.9	97.3 ± 5.0	96.5 ± 5.3	96.9 ± 5.1
BPFS	67.7 ± 25.6	97.6 ± 3.1	92.2 ± 11.4	94.5 ± 7.8

Performances in assessing HS and TO events in US are reported in Table 2, in terms of MAE, precision, recall, and F1-score. LE_5 provides the best MAE in HS and TO identification in US, in terms of both mean and SD (21.6 ± 7.0 ms and 38.1 ± 15.2 ms, respectively). Even LE_{10}, LE_{15}, and $RMSS_{100}$ are able to keep HS-MAE value < 30 ms, but they fail in keeping TO-MAE value < 40 ms. All *LE* and *RMSS* approaches are able to maintain precision, recall and F1-score > 98% for HS and > 97% for TO. *FWRS* and *BPFS* approaches supply the worst performances. Average performances in every subject are reported in supplementary material 1-8.

All experiments run on a machine equipped with a 2,6 GHz Intel Core i7 processor, 16 GB RAM. The best performing signal pre-processing pipeline (LE_5) required approximately 70 milliseconds in average to process a 1-second signal. It then took around 0.2 milliseconds for the neural network to process and predict gait events for a single pre-processed signal window (20 samples). In conclusion, the total processing time sums up to 90 milliseconds to predict TO and HS events for a 1 second walk. In the present experiments, the network training time over 22 training subjects (one single fold) ranges from approximatively 30 minutes, when the simpler experimental protocol is adopted (a single EMG signal per leg, two in total) to approximatively 60 minutes, when all the four EMG signals per leg (eight in total) are used. However, we also note that the network training has to be done only once, then the trained network can be applied as-is to predict TO and HS in unseen subjects.

4.2 Reduction of experimental set-up

Since it turned out to be the best-performing processing technique, LE_5 has been used to perform the analysis of the reduction of experimental set-up. Average classification accuracies as a result of different experimental set-ups are reported in Table 3, for

Learned-test set (LS-test) and Unlearned set (US). The full protocol (reference) provides the best classification accuracy (94.8 ± 0.2% for LS-test and 93.4 ± 2.3% for US). In the distal-leg approach, a significant ($p < 0.05$) decrease of 2 percentage points of classification accuracy is detected for both LS-test and US, compared to the reference. However, accuracy is still widely > 90%. The gap from the reference further increases ($p < 0.05$), considering the single muscles (GL and TA) and the proximal-leg approach (Table 3). Performances in assessing HS and TO events in US are reported in Table 4, in terms of MAE, precision, recall, and F1-score. The full protocol supplies the best MAE in HS and TO identification in US, in terms of both mean and SD (21.6 ± 7.0 ms and 38.1 ± 15.2 ms, respectively) and the best F1-score (99.0 ± 1.7% and 98.4 ± 2.4%). A significant worsening in HS-MAE (≈ + 10 ms, $p < 0.05$) is detected in prediction of distal-leg and GL approaches.

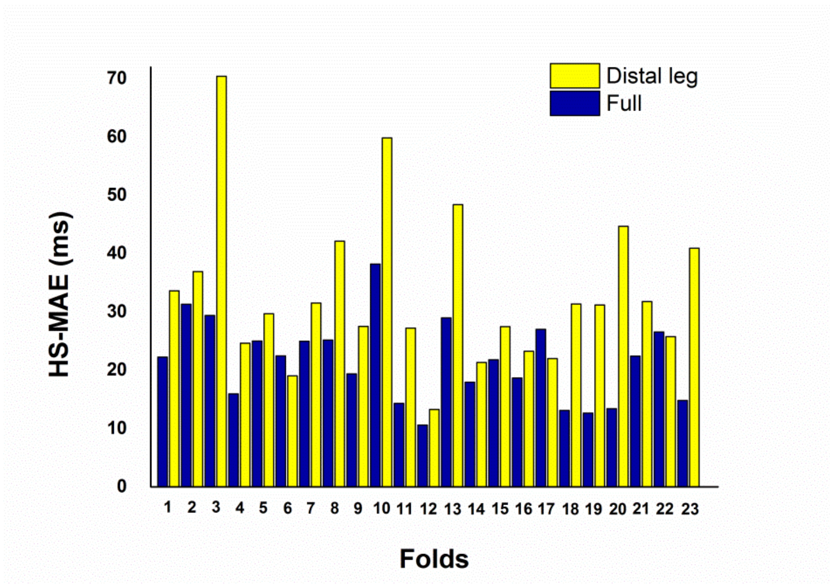
Table 3. Mean classification accuracy as result of different experimental set-ups

	Mean classification accuracy (%)	
	LS-test	**US**
Full	94.8 ± 0.2	93.4 ± 2.3
Proximal leg	84.6 ± 0.9	79.9 ± 6.1
Distal leg	92.6 ± 0.3	91.4 ± 2.6
GL	88.4 ± 0.5	89.1 ± 3.6
TA	86.9 ± 0.4	84.6 ± 6.9

Table 4. MAE (mean absolute error), precision, recall, and F1-score as result of different experimental set-ups for Heel Strike (HS) and Toe Off (TO) prediction in US.

Mean prediction performances				
Heel Strike (HS)	**MAE (ms)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Full	21.6 ± 7.0	99.7 ± 0.6	98.5 ± 3.0	99.0 ± 1.7
Proximal leg	52.9 ± 23.8	96.3 ± 5.0	90.0 ± 9.3	92.8 ± 6.5
Distal leg	33.2 ± 13.1	99.5 ± 0.7	97.3 ± 4.6	98.3 ± 2.7
GL	33.0 ± 12.0	99.7 ± 0.4	96.5 ± 6.7	98.0 ± 4.1
TA	53.6 ± 38.4	95.5 ± 6.1	83.8 ± 15.1	88.7 ± 10.9
Toe Off (TO)	**MAE (ms)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Full	38.1 ± 15.2	99.1 ± 1.5	97.9 ± 3.6	98.4 ± 2.4
Proximal leg	71.2 ± 24.4	92.3 ± 11.0	86.1 ± 11.9	88.9 ± 10.7
Distal leg	45.1 ± 18.7	98.6 ± 2.0	96.4 ± 5.0	97.5 ± 3.3
GL	64.6 ± 25.8	98.9 ± 1.5	95.7 ± 6.7	97.1 ± 4.1
TA	49.0 ± 16.6	95.7 ± 5.7	83.9 ± 15.5	88.8 ± 11.1

A further increase of MAE (≈ + 20 ms, $p < 0.05$) and decrease of F1-score (from 6% to 10%) were predicted by the other two approaches. TO-MAE worsens in prediction of distal-leg and TA approaches (≈ 7 and 11 ms, respectively), even if not significantly ($p > 0.05$). A concomitant decrease of F1-score is detected (≈ -1%). Further remarkable worsening of both parameters was reported for the other two approaches. Figure 1 shows a direct comparison between the accuracy provided by distal-leg (yellow bars) vs. reference set-up (full, blue bars) in each fold. Average performances in every subject are reported in supplementary material 9-12.



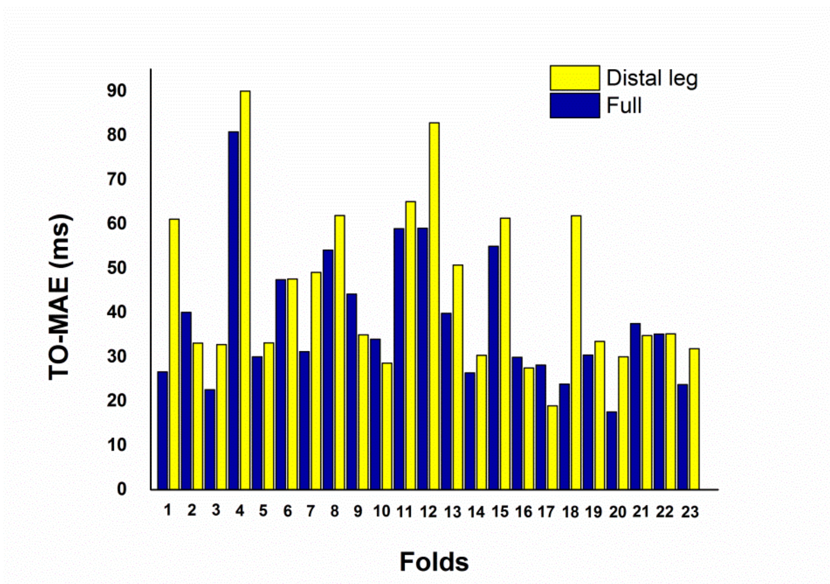


Fig. 1. Direct comparison of MAE provided in each fold by full set-up (dark blue bars) vs. distal-leg set-up (yellow bars) for HS (upper panel) and TO (lower panel) predictions.

5. Discussion

The present group of researchers recently proposed a neural-network-based approach for classifying stance vs. swing and assessing temporal gait events from electromyographic signals [11]. A twofold objective is pursued in the present study, i.e.

to test the influence on the performance of the above-mentioned approach of: 1) different pre-processing of sEMG signal; 2) reduction of the number of sEMG probes included in the experimental set-up. The approach described in [11] was chosen as reference model, because to our knowledge it is still outperforming all similar studies in terms of HS and TO prediction [6-10]. Foot-switch signal was adopted as the ground truth, since it represents the gold standard in gait segmentation [21-23].

5.1 Signal pre-processing

The first step was to test if the change of low-pass cut-off frequency for extracting the envelope could affect classification and/or prediction performances. Average results show that classification accuracy (Table 1) and prediction MAE (Table 2) gradually worsen with concomitantly increasing cut-off-frequency value (starting from the reference value of 5 Hz), in terms of both mean value and SD. These results clearly show that the performances of the classifier are affected by the choice of the cut-off-frequency and 5 Hz is the best value for the goal we set. This suggests that when the envelope of sEMG signal is used, the cut-off-frequency value should be carefully evaluated in relation to the adopted methodology and pursued aim, in order to avoid estimation bias, as shown for co-contraction assessment in [16].

The second step was to compare the performances of the classifier after feeding the neural network with sEMG signal filtered in different ways. The simplest filter analyzed is *BPFS*, because it is necessary for removing low-frequency motion artefacts and high-frequency noise from the signal. This approach returns the worst mean and single-subject (see supplementary material 8) classification accuracy among the approaches considered in the present study, especially for the unseen subjects (mean ± SD = 84.0 ± 3.7%, Table 1). The worst performances are provided also in the assessment of gait events (Table 2).

BPFS is the only approach not performing the rectification of the signal. Thus, these findings indicate that the full-wave rectification is strongly recommended in processing the signal to feed the neural network. However, the full-wave rectification alone does not seem to be enough. *FWRS* approach truly improves *BPFS* one, but accuracy is still < 90%. Moreover, the performances are far from the ones provided by more refined processing approaches, such as *LE* and *RMSS* (Table 1 and 2). All *LE* and *RMSS* approaches, indeed, report a mean classification accuracy > 90% (Table 1) and keep mean precision, recall and F1-score > 98% for HS and > 97% for TO (Table 2). Furthermore, the best performing *LE* approach (LE_5) outperforms also the *RMSS* approaches, above all in terms of average classification accuracy (≈ 95% in learned subjects and > 93% in unseen subjects), HS-MAE (21.6 ± 7.0 ms vs. 28.1 ± 9.6 ms provided by the best performing *RMSS* approach), and TO-MAE (38.1 ± 15.2 ms vs. 54.1 ± 29.5 ms). About *RMSS*, the different durations of sliding-window do not seem to influence the classifier performance.

In the end, present results confirm that the choice of the sEMG processing actually affects the classification/prediction performances, as expected [15-17]. Moreover, the present study succeeds in identifying the linear envelope (cut-off frequency 5 Hz) as the sEMG-processing type which provides the best performance of the neural network in terms of both classification accuracy and gait-event-prediction, among the four widely-used approaches analyzed in the present study. This methodological finding, reported here for the first time, is very useful information for improving the precision of the clinical test by means of the most adequate processing of the signal. It seems especially valuable for those clinical conditions (such as neurological disorders) where elevated precision of predictions is fundamental to properly identify subject recovery during follow-up.

5.2 Reduction of experimental set-up

Since LE_5 is resulted being the best-performing approach for the sEMG-signal processing, it was used to run the analysis of the reduction of experimental set-up, i.e. the number of sEMG probes. Besides the full protocol [11], four reduced experimental set-ups are considered in the present paper, in order to test the influence of protocol simplification on classification/prediction performances. The first step was to test if the reduction from four to two sEMG sensors per leg could provide classification/prediction results consistent with those provided by the full set-up. Two attempts were made, using signals from a couple of sensors applied to the same leg segment (proximal or distal), one in the front and one in the back. Table 3 shows as mean classification accuracy provided by the proximal-leg set-up clearly deteriorates compared to the full set-up, falling below 85% in learned subjects and below 80% in unseen ones. This is also more evident by analysing each single subject, as reported in supplementary material 9. The proximal-leg-based reduction of the number of sensors strongly affects also MAE, precision, recall, and F1-score, especially in TO prediction (MAE = 71.2 ± 24.4 ms and F1-score < 90%, Table 4). The matter is different for the distal-leg set-up. Although a decrease of mean classification accuracy is still detected, it amounts to only 2 percentage points in both learned (92.6 ± 0.3%) and unseen subjects (91.4 ± 2.6%). Moreover, precision, recall, and F1-score remain practically unaltered. The mean increase of MAE compared to the full set-up (+ 11.6 ms for HS; + 7.0 ms for TO, Table 4) is the price to pay for using only two probes per leg. For allowing a more detailed evaluation, MAEs provided by the two set-ups in each single subject are compared in Fig. 1. These findings show as the distal-leg set-up clearly outperforms the proximal-leg one in terms of all performance parameters. To our knowledge, only Nazmi at al. [10] provided neural-network prediction of gait events using only two sEMG probes per leg on distal-leg muscles (tibialis anterior and gastrocnemius medialis). They achieved, for unseen subjects, a mean classification accuracy of 77% and MAE of 35 ms and 49 ms in assessing HS and TO, respectively. Compared to those, the present distal-leg-set-up results appear promising, considering also that in present study F1-score is around 98%, while in [10] this information is not reported. It is worth mentioning that in [10] HS and TO predictions are computed in only 5 seconds of a single subject, whereas in the present study an extensive evaluation is performed, predicting HS and TO in a 5-minute walking of 23 different subjects (with leave-one-out cross validation).

The second step was to test the effect of a further reduction to a single muscle of experimental set-up. Considering the promising results achieved, two further attempts were made, using signals from one sensor applied to a single muscle, in the front (TA) or in the back (GL) of the distal leg. The results shown in Table 3 highlight that phase classification based on a single sEMG signal leads to a deterioration of mean accuracy, compared to both full and distal-leg set-ups. This is true for both TA (- 8% in learned and - 7% in unseen subjects, compared to the full protocol) and GL set-ups (- 6% in learned and - 4% in unseen subjects), although GL set-up achieves better accuracy, getting close to 90% in unseen subjects. However, classification accuracies are still better than the ones provided by the proximal-leg set-up and in [10]. TA and GL set-ups work differently in HS and TO prediction. Compared to full set-up, mean precision, recall, and F1-score remain practically unaltered for GL set-up. Prediction of TA set-up is instead affected by a strong decrease of mean recall value (-14% and -12% compared to full and GL set-ups, respectively). This means that a high number of false positive detection of HS and TO affects the prediction. Consequently, a concomitant deterioration of F1 score is observed (-10% and -9% compared to full and GL set-ups, respectively). HS prediction provided by GL set-up presents a mean increase of MAE compared to the full set-up, but achieves the same value provided by distal-leg set-up (33.0 ± 12.0 ms vs. 33.2 ± 13.1 ms). Furthermore, mean HS-MAE is still comparable with the one reported in [10], using two distal-leg muscles. TA set-up reports a significant growth of mean HS-MAE compared to full (+33 ms), distal-leg (+20 ms), and GL (+20 ms) set-ups. These findings seem to indicate that between GL and TA signals, only GL-signal plays a fundamental role in prediction of heel strike. TO prediction is less accurate in GL-set-up (+26 ms and +19 ms of mean MAE compared to full and distal-leg set-ups, respectively). Larger MAEs in TO prediction were foreseen, since it was explained that it is more challenging to assess TO than HS [10,24]. On average, TO-MAE is lower for TA set-up (49.0 ± 16.6%, comparable with distal-leg set up). However, as already mentioned, it is associate to low performances in terms of classification accuracy (84.6 ± 6.9%), recall (83.9 ± 15.5%), and F1-score (88.8 ± 11.1%). Thus, in our opinion TA-set-up-based prediction should be considered not reliable and the desirable simplification of experimental set-up (one single sensor) should involve only the GL set-up. In this case, this simplification would be paid with a deterioration of TO (not HS) prediction. This could be a good compromise for tasks such as stride recognition, stride-time computation, identification of toe walking, and so on, where only HS event is involved.

In the end, present findings indicate that the reduction of the complexity of the experimental sEMG-protocol (i. e. decreased number of sEMG sensors) affects the performances especially in terms of gait-event-prediction parameters, as expected [10]. Moreover, the present study succeeds in the goal of providing for the first time a quantification of the progressive deterioration of classification/prediction performances with the reduction of the number of sensors used. This could be very useful in clinics to the aim of choosing the most suitable approach, balancing technical performances, patient comfort, and clinical needs. Since a simplification of experimental set-up is always desirable, the present study proposes the distal set-up (consisting of two sensors over TA and GL per leg) as a suitable alternative to the full protocol in those circumstances where limiting time consumption and patient discomfort is a primary issue. The price to pay for this simplification is essentially a worsening of HS and TO prediction (about 10 ms, on average). A further reduction of experimental set-up to a single muscle seems to be feasible without a further deterioration of performances only if GL is chosen as the reference muscle and for computation where only heel-strike events are involved.

6. Conclusions

The present study shows that both the sEMG-processing type and the reduction of sEMG-protocol complexity actually affect the performances of neural-network-based classification of gait phases and assessment of temporal gait events. A further novel contribution is to provide also a reliable quantification of this performance deterioration. The quantitative knowledge of the consequences of the reduction of the number of sensors in terms of classification/prediction accuracy could be very useful in clinics to drive the choice of the most suitable experimental set-up for gait analysis, able to balance the need of handling patient comfort and limiting time consumption with the necessity of maintaining an elevate precision of test results. Higher precision in gait event prediction is increasingly requested in clinics, especially in those pathologies where one of the gait phases could be strongly reduced (neurological disorders). The present study provides also the information about the most suitable sEMG-processing type (linear envelope with a cut-off frequency = 5 Hz) to satisfy this necessity.

Four acknowledged and widely-used approaches to process EMG signals were included in the present comparative analysis. Future development could be designed to involve more advanced signal-processing techniques in frequency or time/frequency domain, such as Fourier transform or wavelet transform. Moreover, the potential influence of gait velocity could be also taken into account. This would be an intriguing further direction, as EMG envelopes show adaptations to different gait velocities.

References

1. Perry J. Gait Analysis - Normal and Pathological Function. USA: Slack Inc, 1992.
2. Bovi G, Rabuffetti M, Mazzoleni P, Ferrarin M. A multiple-task gait analysis approach: Kinematic,kinetic and EMG reference data for healthy young and adult subjects. Gait Posture. 2011:33;6–13.
3. Caldas R, Mundt M, Potthast W, Buarque de Lima Neto F, Markert B. A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms. Gait Posture. 2017:57;204–210.
4. Taborri J, Palermo E, Rossi S, Cappa P. Gait Partitioning Methods: A Systematic Review. Sensors (Basel) 2016;16(1):66. doi: 10.3390/s16010066
5. Di Nardo F, Morbidoni C, Cucchiarelli A, Fioretti S. Recognition of Gait Phases with a Single Knee Electrogoniometer: A Deep Learning Approach . Electronics 2020:9(2);355. https://doi.org/10.3390/electronics9020355
6. Meng M, She Q, Gao Y, Luo Z. EMG signals based gait phases recognition using hidden Markov models. In Proceedings of the 2010 IEEE International Conference on Information and Automation, Harbin, China, 20–23 June 2010; pp. 852–856.

7. Joshi CD, Lahiri U, Thakor NV. Classification of gait phases from lower limb EMG: Application toexoskeleton orthosis. In Proceedings of the 2013 IEEE Point-of-Care Healthcare Technologies (PHT),Bangalore, India, 16–18 January 2013:228–231.
8. Nazmi N, Abdul Rahman M, Ariff MHM, Ahmad S. Generalization of ANN Model in Classifying Stance and Swing Phases of Gait using EMG Signals. In Proceedings of the IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES 2018), Sarawak – Kuching, Malaysia 3-6 December. 2018: 461-466.
9. Ziegier J, Gattringer H, Mueller A. Classification of Gait Phases Based on Bilateral EMG Data Using Support Vector Machines. In Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics, Enschede, The Netherlands, 26–29 August. 2018:978–983.
10. Nazmi N, Abdul Rahman M, Yamamoto SI, Ahmad S. Walking gait event detection based on electromyography signals using artificial neural network. Biomed Signal Process Control. 2019;47:334–343.
11. Morbidoni C, Cucchiarelli A, Fioretti S, Di Nardo F. A Deep Learning Approach to EMG-Based Classification of Gait Phases during Level Ground Walking. Electronics 2019:8(8);894. https://doi.org/10.3390/electronics8080894
12. Morbidoni C, Principi L, Mascia G, Strazza A, Verdini F, Cucchiarelli A, Di Nardo F.Gait Phase Classification from Surface EMG Signals Using Neural Networks. In: Henriques J, Neves N, de Carvalho P (eds). XV Mediterranean Conference on Medical and Biological Engineering and Computing - MEDICON 2019.
13. Kamruzzaman J, Begg RK. Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait. IEEE Trans Biomed Eng. 2006;53(12 Pt 1):2479-90.

14. Tang Z, Zhang K, Sun S, Gao Z, Zhang L, Yang Z. An upper-limb power-assist exoskeleton using proportional myoelectric control. Sensors (Basel). 2014;10;14(4):6677-94.
15. Rosa MC, Marques A, Demain S, Metcalf CD, Rodrigues J. Methodologies to assess muscle co-contraction during gait in people with neurological impairment – a systematic literature review. J Electromyogr Kinesiol. 2014; 24(2):179–191.
16. Rinaldi M, D'Anna C, Schmid M, Conforto S. 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. 2018;43:184-192. doi: 10.1016/j.jelekin.2018.10.007.
17. Caramia C, De Marchis C, Schmid M. Optimizing the Scale of a Wavelet-Based Method for the Detection of Gait Events from a Waist-Mounted Accelerometer under Different Walking Speeds. Sensors (Basel). 2019;19(8):1869.
18. Winter DA, Yack HJ. EMG profiles during normal human walking: stride-to-stride and inter-subject variability. Electroencephalogr Clin Neurophysiol. 1987;67(5):402-411.
19. Hermens HJ, Freriks B, Merletti R, et al. European recommendations for surface electromyography, SENIAM. Enschede (NL): Roessingh Research and Development 1999, 8.
20. Di Nardo F, Mengarelli A, Maranesi E, Burattini L, Fioretti, S. Gender differences in the myoelectric activity of lower limb muscles in young healthy subjects during walking. Biomed Signal Process Control. 2015;19:14–22.
21. Agostini V, Balestra G, Knaflitz M. Segmentation and Classification of Gait Cycles, IEEE Trans. Neural Syst. Rehabil. Eng. 2014;22(5):946-52.

22. Taborri J, Palermo E, Rossi S, Cappa P. Gait partitioning methods: A systematic review. Sensors. 2016;16:66.

23. Winiarski S, Rutkowska-Kucharska A. Estimated ground reaction force in normal and pathological gait. Acta Bioeng. Biomech. 2009;11:53–60.

24. Khandelwal S, Wickstrasm N. Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the MAREA gait database. Gait Posture. 2017;51:84–90.