The reliability of objective fatigue measures in Multiple Sclerosis Patients

The reliability of objective fatigue measures in Multiple Sclerosis Patients

Francesca Farolfi[1], Federica Miola[1], Gabriella Olmo[1], *Senior Member, IEEE,* Stefano Giorgi[2], Francesca Sperli[2], Alessia Di Sapio[2], Antonio Bertolotto[2]

1: Dept. of Control and Computing Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, Italy.

2: 2nd Neurology Dept., Regional Multiple Sclerosis Centre, San Luigi Gonzaga Hospital, Orbassano, Italy

***Abstract*— Fatigue is one of the most distressing symptoms of Multiple Sclerosis, impairing quality of life, work performance and social interactions. Moreover, it is difficult to objectively assess, and is often evaluated via subjective questionnaire; however, objective metrics are highly desirable.**

Objective of this paper is to implement a differential fatigue measure based on evoked potentials, and assess its reliability and coherence with walking tests and subjective questionnaires.

Method: the Regional Multiple Sclerosis Reference Centre, San Luigi Gonzaga Hospital, Italy, is carrying out tests in order to assess the effectiveness of 4-amino-pyridine for fatigue relief. This work takes as input evoked potential waveforms and implements an algorithm to perform suitable signal processing and provide a fatigue index and a related reliability metric. This information is put in relationship with the walking test results and the subjective fatigue scores.

Preliminary results reveal that fatigue measures based on evoked potentials, subject to proper data processing, are not always coherent with subjective questionnaires and walking tests. This may be due to the fact that fatigue cannot be reduced to a mere muscular/conduction problem, and the walking tests are heavily conditioned by the disability degree. Moreover, the reliability of such measures carries not trivial information that should be carefully considered.

As a conclusion, evoked potentials and gait analysis represent a good complement of subjective questionnaires as for fatigue assessment.

The significance of our work lies in the fact that reliable fatigue measures can help improving the patients' quality of life, allowing assessment of the therapy effectiveness and posology.

I. Introduction

In Multiple Sclerosis (MS), a disimmune reaction against the myelin basic protein provokes inflammation of the myelin sheaths around axons of the Central Nervous System (CNS) [1]. Defective remyelination triggers subsequent neurodegeneration; hence, the natural course of the disease consists in an early inflammatory stage, followed by a chronic phase where plaques are completely evolved in scars.

MS is mainly diagnosed in young adults between 20 and 40 years [2]. It is considered a multifactorial disease, and a number of risk factors are recognized or suspected, such as environment and ethnicity (latitude, temperate climate, Caucasian origin), toxic agents, low levels of vitamin D, exposures to infective agents, genetic predisposition. At present, many disease-modifying therapies are indeed available, which alleviate the long-term disability outcomes, especially if started early in the course of the disease.

MS signs and symptoms are variable among subjects, due to the different affected areas of the CNS. Common symptoms, especially in early stages, are impaired visual acuity, non-controlled eye movements and diplopy; disorders of sensibility: persistent tingling, numbness of limb, loss of touch sensitivity, problems in distinguishing cold and warm, loss of muscular strength. On the other hand, the so-called *hidden symptoms* encompass cognitive dysfunction, depression, anxiety, and fatigue.

Fatigue is one of the most distressing and common symptoms of people with multiple sclerosis (PwMS), with a prevalence of 75%-90% [2]. The Multiple Sclerosis Council for Clinical Practice defines this phenomenon as *a subjective lack of physical and/or mental energy, perceived by the individual or caregiver to interfere with usual and desired activities*; fatigue is classified as chronic if it is present, for any amount of time, on 50% of the days for more than six weeks [3].

Some therapeutic strategies can improve quality of life. For example, fampridine, a slow-release formula of 4-aminopyridine, is assumed to have a positive impact on fatigue, even though not all PwMS report this effect [4], [5].

Fatigue is usually evaluated by means of subjective scales [6], i.e. questionnaires filled in by the PwMS themselves and reporting his/her subjective perception on several items. However, clinicians agree that objective measures may help quantifying the phenomenon; in particular, reliable differential measures could allow one to appreciate subtle variations in fatigue, so as to monitor the symptom evolution over time and the efficacy of therapies.

A possible method to objectively evaluate fatigue makes use of motor evoked potentials [7], which allow one to study both central and peripheral nervous conduction and to appreciate variations. However, these measures, besides being expensive and burdensome for patients, are subject to unavoidable error sources that make their interpretation not trivial. Moreover, they only reflect muscular fatigue, which is only one aspect of this complex phenomenon.

Another way to objectively measure fatigue is represented by walking tests, such as the 6-minute- or the timed 25-feet walking tests [8], [9], [10]. Fatigue can be put in relationship

with the distance covered in a fixed amount of time (or, conversely, the time employed to walk a fixed distance). Again, walking tests mainly represent muscular fatigue; moreover, in advanced stages of the pathology, PwMS can be unable to perform walking tests, or results may be unreliable.

From this brief discussion, it can be understood that possible sources of information related to fatigue are not exhaustive *per se*, and may be affected by several unavoidable sources of instability and confounding factors. This paper finds its motivations in this context, and the main contributions are manifold.

- Starting from a protocol implemented in the clinical practice [11], [12], [13], [14], we provide an algorithm to extract an automatic differential fatigue score from motor evoked potential waveforms.
- In order to manage unavoidable instability sources, we propose a reliability metric to be associated with these measures.
- We address walking test data and subjective questionnaires, define a fatigue measurement index for each of them, and verify whether these indices yield information coherent to motor evoked potentials.

In more detail, this study exploits measures performed at the Regional Multiple Sclerosis Reference Centre, 2nd Neurology Department of San Luigi Gonzaga Hospital, Orbassano, Italy, where a clinical trial is being performed to assess the responsiveness to fampridine for fatigue relief. As part of this trial, eligible PwMS have been clinically evaluated by expert neurologists and submitted to evoked potentials and walking tests. Then, after two-week fampridine administration, the tests have been repeated in order to verify whether an objective variation in fatigue could be appreciated. We have employed the recorded evoked potential waveforms to automatically implement a differential fatigue score, following the principles described in [12], [13]. We have worked out a *reliability metric* assess the validity of evoked potentials for this purpose. We have employed low-cost, easy-to-use instrumentation, such as a smartphone to register gait data on the same patients enrolled in this trial. The results of walking tests and subjective fatigue questionnaires are used to work out a *motor score* and a *subjective score*. These different sources of information have been compared with each other and with those manually worked out by clinicians, and some preliminary conclusions have been drawn.

The rest of this paper is organized as follows. In Sect. II the complex phenomenon of fatigue is described in some detail, and focus is put on methods for state-of-the-art subjective and objective fatigue evaluation. In Sect. III the proposed approach is described. In Sect IV preliminary results are presented, and in Sect. V conclusions are drawn.

II. Theory: measuring fatigue, state of the art

Despite its high prevalence, the physio-pathology of fatigue is still not completely clear [15]. *Primary fatigue*, a mental and physical weakness not alleviated by sleep or rest, worsened by high ambient temperature or when a single muscle group is intensively used for a short period of time, finds its causal sources in the demyelinating process at central level, which implies an irreversible reduction in glucose metabolism. On the other hand, *secondary fatigue* is a consequence of MS symptoms or side effects of drug therapies [16]. It is known that 50% of people with daytime fatigue also present nocturnal sleep disturbances. Many PwMS may develop depression and/or anxiety during their lifetime, with related sensations of exhaustion or tiredness. Hormonal imbalance (e.g. thyroid dysfunction, possibly induced by long-term interferon-beta therapy or alemtuzumab) could play a role in this symptom, along with anaemia, vitamin D deficiency and other related health problems. Moreover, many drugs could affect fatigue occurrence: interferon-beta itself, immunosuppressant drugs, muscle relaxants, antidepressants, hypnotics and benzodiazepines. Hence, multifaceted measures are likely to be necessary to represent such a complex phenomenon.

A. Subjective Fatigue Indicators

The state-of-the-art assessment of fatigue is based on self-reports. Some of the most employed ones are [6]:

- *Fatigue Severity Scale* (FSS). The PwMS is asked to comment a list of nine items related to her/his fatigue perception, with a score between 1 (strongly disagree) and 7 (strongly agree). It ranges between 9 and 63 (worst case).
- *Modified Fatigue Impact Scale* (MFIS). It is composed of 21 items about sensations felt during daily activities, each of which is scored depending on its frequency in the past 4 weeks (never, rarely, sometimes, often, very often). It ranges between 0 and 84 (worst case).
- *Visual Analogue Scale* (VAS). It consists of a line with a starting point marked with 0 (absence of fatigue) and a ending point marked with 10. The PwMS has to indicate a point related to the intensity he/she feels.

B. Objective Fatigue Estimation

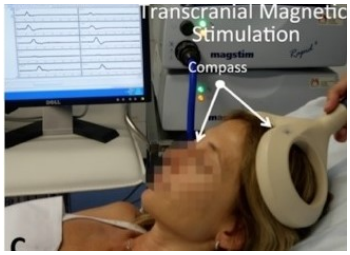
Any objective metric will necessarily take into account mainly physical aspects of fatigue; nevertheless, it can be a valuable tool in integrating subjective scales with quantifiable information. In literature, papers are reported which address MS fatigue evaluation by means of motor evoked potentials and gait analysis [17], [18], [19], [15].

Motor Evoked Potentials. Motor Evoked Potentials (MEPs) [7] represent the CNS electrical response to external stimuli, and are fundamental in the assessment of conduction in motor neuron diseases, spondylotic myelopathy, stroke, ataxia, spastic paraplegias and MS [12]. They are recorded at muscular level after transcranial magnetic stimulation. In [12] a double cone coil is placed in correspondence of the deep cortical regions related to the lower limb motion. A transient magnetic field induces electric currents in the brain, and yields direct bilateral stimulation of the cortical leg motor area. At least five magnetic stimuli are given, and averaged to achieve the signal to be analyzed. Then, MEPs are registered in a number of distal and proximal muscles. The Central Motor Conduction Times (CMCT) are obtained as the subtraction of the

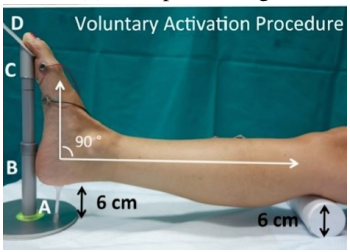
peripheral conduction time from the latency of corresponding MEPs. The peripheral conduction time is evaluated via root-Compound Motor Action Potentials (rCMAPs), elicited in the same districts via high voltage electrical stimulation of lumbo-sacral nerve roots at their origin from the spinal cord [11]. Stimuli are delivered with an intensity that guarantees responses with amplitude from 50 to 200 μV in almost all recording sites. In this way, corticospinal pathways excitability and integrity can be assessed using MEP and rCMAP areas and latencies.

Recent studies have revealed that a decrease in the area under the MEP/rCMAP curves, as well as increased MEP/rCMAP latencies, are significant as for fatigue evaluation [20], [12], [13]. Notably, the use of areas is challenging because of their variability in serial recordings, due to temperature variations, shift in the stimulation/recording sites, general patient conditions. In [12], [13] a protocol is proposed to reduce this variability. A standardized voluntary muscle activation is employed, and modifiable causes of variability are controlled to the maximum possible extent. For example, the authors suggest that a multiple electrode array is placed over the dorso-lumbar tract, and the site is marked by a small round tattoo, similarly to radiotherapy protocols (Fig. 1). Two electrical stimuli are given to the PwMS to record peripheral muscle activation signals, and the last epoch is selected for the analysis.

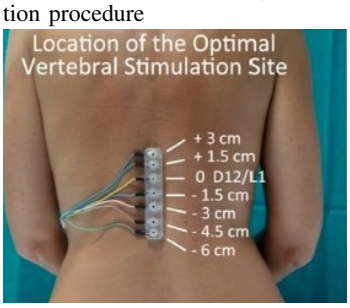
If this protocol is properly implemented, residual differences between areas and latencies, measured on the same PwMS in different recording sessions, should reflect actual modifications in nerve conduction. Nevertheless, as discussed in Sect. III, a number of unavoidable error sources impair the interpretation of such data. This is the reason why we address data pre-processing and automatic evaluation of evoked potentials, and work our a reliability index for such measures.



(a) Coil positioning



(b) Detail of the voluntary activation procedure



(c) Electrode array placement for peripheral stimulation

Fig. 1: Details of protocol for muscle activation (from [13], with permission of the authors)

Gait analysis. A progression in fatigue may be estimated by means of the distance walked by the subject in a fixed amount of time, and significant differences can be appreciated between PwMS and healthy controls [21]. Commonly performed tests are:

- Timed 25-Feet Walking Test (T-25FWT). The PwMS is required to walk 25 feet as quickly as possible. The time spent to cover this distance is then annotated.
- 6-Minute Walking Test (6MWT). The distance walked in 6 minutes is annotated. In case the PwMS is unable to walk for 6 minutes, a possible variant is the 2-Minute Walking Test (2MWT).

Moreover, subtle alterations of walking paths may measure motor impairment in many neurodegenerative disorders [9], [10]. The attitude of subjects to walk slowly, take short steps, increase the time of foot support, is known to be correlated with the degree of disability. In [17], secondary gait compensations occurring in PwMS, such as head and pelvis movement, are measured via a wearable tri-axial accelerometer and put in relationship with gait stability. Twelve healthy participants and 12 PwMS performed the 6MWT, and measures of gait compensation, mobility, variability, asymmetry, stability and fatigue were gathered. Compared to controls, PwMS exhibit greater vertical asymmetry in head and pelvic movements; increased compensatory movements were correlated with higher disability. However, no clear relationship with fatigue was put into evidence. In [18] one hundred relapsing-remitting PwMS, divided into mild and moderate groups depending on the neurological impairment level, were asked to walk on an electronic walkway. Gait velocity, cadence, step/stride length, swing/stance time and several other parameters were measured. Fatigue turned out to be moderately correlated with a few of such gait parameters. On the other hand, in [19], the authors revealed a significant negative correlation between velocity, cadence and stride length (evaluated on twenty-four PwMS and 19 healthy subjects) and reported fatigue. In [22] one hundred and eighty-nine older adults were assessed for 16 gait parameters while executing the 2MWT on an instrumented walkway. Age, executive function, attention, balance, and physical fatigue turned out to be independently associated with some of these gait parameters.

From these considerations, it turns clear that gait parameters are not straightforwardly related to MS fatigue. Nevertheless, they may help in quantifying it, if associated with different measures. It is worth noticing that the gold standard method for gait analysis is represented by motion capture systems equipped with infrared cameras and retro-reflective markers

[23]. However, these systems are very expensive and may not be available in MS centers. In this work, we measure inertial gait data using simple smartphones. Even though we only address 6MWT and T-25FWT measures, the validation and use of more subtle gait parameters for fatigue evaluation using low-cost instrumentation is left for future developments.

Fampridine and Fatigue. Fampridine [4], [15] is used in the treatments of motor symptoms in PwMS. The active substance is 4-amino-pyridine, a potassium voltage-dependent channel blocker. Numerous studies have shown that K-channels are over-expressed in demyelinated neuronal membranes, hence are responsible of many impairments at motor level. The ability of 4-amino-pyridine to block K-channels may help in the re-establishment of conduction along axons. Fampridine is also thought to improve MS motor fatigue; however, as this effect is not reported by all PwMS, the available data warrant further verification to recommend this drug for the treatment of MS fatigue [15]. Experiments related to assess this aspect can also be considered as useful case studies for the more general problem of MS fatigue evaluation.

In [5], 112 adult PwMS with walking disability (EDSS score[1] 4-7) were administered fampridine 10 mg twice daily. As potential benefits of fampridine can be appreciated after a 15-day administration, the PwMS were evaluated before and after 2 weeks of treatment by means of the 2MWT. The distances covered each 30 s were noted. A fatigability index was defined in terms of the relative difference between the speed over the first and the last 30 s of the test. Walking capacities were subjectively assessed using the MSWS-12[2], and an improvement of at least 15% was considered clinically meaningful. Other effectiveness parameters (walking endurance in the 2MWT, subjective self-perceived walking ability on the MSWS-12) were considered for the definition of fampridine responder, and the same 15% threshold was arbitrarily applied. According to this composite criterion, PwMS were considered responders if they displayed an improvement of 15% or more in at least one test. However, the responders and non-responders turned out not to exhibit significant differences in terms of baseline demographic and disease characteristics.

In [11], [12], [13], [14] evoked potentials are employed to evaluate fampridine effectiveness for fatigue relief. The *inter-trial variability (ITV)* metrics, originally defined to assess the stability of repeated measures [13], is used to assess modification in the physical fatigue phenomenon if the tests are performed before and after the 15-day administration of fampridine [14], so yielding the target differential fatigue measures addressed in this paper.

III. Materials and Methods

At present, ten PwMS have been enrolled at the Azienda Ospedaliero Universitaria San Luigi Gonzaga (three women and seven men) with average age 48.4 years (range 24 - 66) and average EDSS score 4.5 (range 1 - 6.5). The study was conducted in accordance with the Declaration of Helsinki; written informed consent was obtained from all the involved subjects, and the study was approved by the local Ethics Committee. After a clinical evaluation, and having filled in the subjective fatigue questionnaires, PwMS have been subject to transcranial magnetic stimulation and high voltage electrical stimulation to record MEPs and rCMAPs. The protocol in [12] has been adopted for both pre- and post-drug tests. The responses of ten leg muscles have been recorded, i.e. (right/left) Vastus Medialis, Vastus Lateralis, Tibialis Anterior, Peroneus Longus, Flexor of Hallucis Brevis. Moreover, participants were asked to perform the T-25FWT and 6MWT, with inertial data being recorded using a Samsung S5 mini smartphone secured to the patient's waist. The time spent and walked distance were also manually annotated by a technician. Then, fampridine was administered for two weeks, and all the tests repeated.

A. Starting point: clinical evaluation of evoked potential

The clinical evaluation of evoked potentials was performed as in [14]. It is based on the manual comparison of MEP/rCMAP latencies and areas as directly output by the measurement instrumentation, composed of a 16-channel bipolar amplifier (BrainAmp ExG, Brain Products GmbH), Digitimer Ltd. stimulators and related software interface [11]. The differential (ITV) metrics defined in [13] have been evaluated as follows.

Given a generic measure M_{pre} (M_{post}) taken before (respectively after) fampridine administration, the related ITV is defined as:

$$ITV_M = \frac{M_{post} - M_{pre}}{0.5(M_{post} + M_{pre})} \cdot 100 \qquad (1)$$

For each of the ten addressed muscles, six ITV values have been evaluated as in [12]. Two of them refer to central latency and area measures; two of them address peripheral latency and area measures; finally, two metrics address mixed measures. In more detail, the following data are obtained:

1) **MEP latencies**, evaluated as the time elapsed between the stimulation and the appearance of the related response, i.e. the time instant when the MEP signal reaches a given percentage of its maximum value (set to 3% as in [12]).
2) **MEP areas**. ITVs for areas are normalized as in [12], to avoid potential bias due to negative correlation between ITV and absolute area values.
3) **rCMAP latencies**, evaluated similarly on peripheral data.
4) **rCMAP areas**, normalized as in [12].
5) **CMCT,** evaluated as the difference between rCMAP and MEP latencies.
6) **Area Ratio (AR),** i.e. the ratio between areas of central and peripheral potentials. AR is recognized as one of the most important neurophysiological indices to detect conduction failures [13].

Each PwMS is then characterized by 60 ITV parameters for each measurement session. The clinicians employ such

[1]The Expanded Disability Status Scale (EDSS) is a clinical scale that measures the level of disability. It ranges from 0, corresponding to normal neurological exam, to 10, which represents the maximum level of disability

[2]The Multiple Sclerosis Walking Scale (MSWS-12) is a PwMS-based scale that evaluates the impact of MS on several aspects of walking.

parameters to work out a global score related to fampridine responsiveness, as follows.

- Each one of the 60 ITVs is compared with a range determined by 5-th and 95-th percentile of its sample distribution, specific for each muscle and evaluated on a cohort of unaffected volunteers and PwMS [12], [14]. A value 1 (0,-1) is assigned if the ITV under examination suggests improved (stationary, impaired) conduction, depending on the distribution percentile interval the ITV value belongs to.
- Each score is summed up over the ten muscles under examination, so as to achieve six indices ranging from -10 (worst case: the ITV at hand denotes impairment on all the ten muscle locations) to 10 (best case: the ITV at hand denotes improvement for all the ten muscle locations).
- If a given score is larger than or equal to four, the PwMS is considered responsive as for that ITV parameter, otherwise she/he is classified as stationary (impaired) for that parameter [13]. In thios way, six thresholded indices are obtained, whose possible values are -1,0,1. These six indices are summed up, achieving a *Global Score* ranging between -6 and 6. If the Global Score is larger than or equal to two, the PwMS is classified as a fampridine responder [14].

The clinicians use the results related to the walking tests as a term of comparisons. If a PwMS, classified as a responder, has exhibited a sensible improvement in his/her walking performance (namely, he/she has walked at least a 30% longer distance after fampridine administration [14]), this is considered as a further confirmation that the PwMS is actually a responder. It must be noticed that, having set a very selective threshold, hardly any PwMS has been considered as a responder based on the walking test results [14].

B. Contribution 1: automatic evaluation of evoked potential global score

The waveforms recorded as described in Sect. III-A have been input to an algorithm devoted to the Global Score automatic evaluation. The algorithm is implemented in `Matlab`, version 2018a for Windows 10.

Signal pre-processing. Central and peripheral evoked potentials are pre-processed by the following steps:

- rectification;
- band-pass filtering using a 10-order Butterworth filter, bandwidth 3 Hz-3 kHz.
- evaluation of the ITV values.

A preliminary analysis of MEP/rCMAP signals pointed out some issues related to instability in the signal acquisition process. A large variation in peak values of the *pre/post signal pairs* (i.e., corresponding signals before and after fampridine administration) may be often appreciated; actually, in about 30% of both MEP and rCMAP signals, the pre/post peak values differ more than twice (Fig. 2). Such a large difference, not meaningful from the physio-pathological point of view, is likely related to unavoidable measure errors. For example, the positions of the stimulation and recording electrodes is very difficult to control, despite the countermeasures included in protocol [12]. Some signals exhibit extremely low peak amplitudes (Fig. 3) with hardly recognizable waveform, suggesting an electrode detachment. A non-canonical signal morphology can occur (see for example Fig. 4), as can be easily quantified using pattern matching methods similar to those applied for ECG analysis [24]. These situations are difficult to control, as the measures, even though conducted using a protocol to limit systematic errors, can hardly be repeated as they are time-consuming and inconvenient for the PwMS. Moreover, the impact of these anomalies in the global score evaluation is not clear.

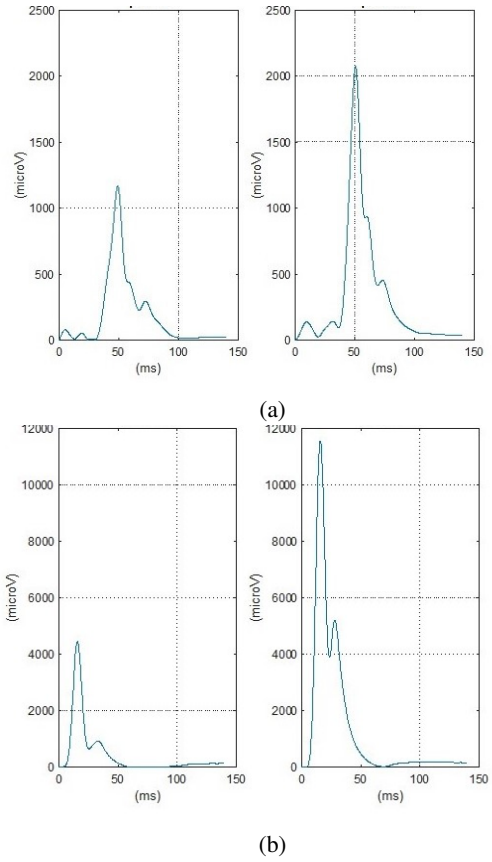


Fig. 2: Peak difference between pre- and post-fampridine signal pairs. (a): MEP. (b): rCMAP.

We have implemented a systematic approach to circumvent such issues.

1) Those signals whose peak value is below 50 μV are not considered, so as not to assign improper significance to meaningless data. In fact, the stimulation hardware is tuned so as to provide responses whose peak amplitude exceeds 50 μm.
2) Those waveform that do not meet proper criteria are discarded *a priori*. Let us focus on the peak variability problem, and let us define

$$P = |P_{pre} - P_{post}| \tag{2}$$

as the difference between the peak values of pre/post signal pairs. We have estimated the P sample distribution for all the 100 available signal pairs. The sample

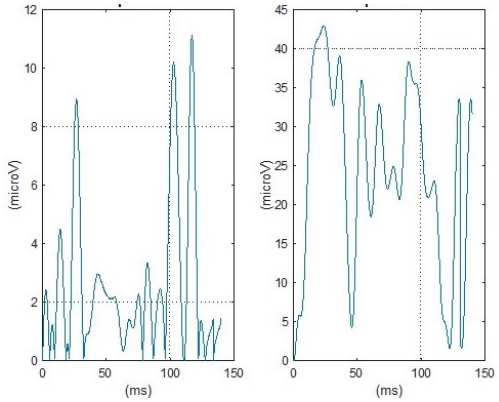


Fig. 3: A signal pair whose peak is below 50 μV.

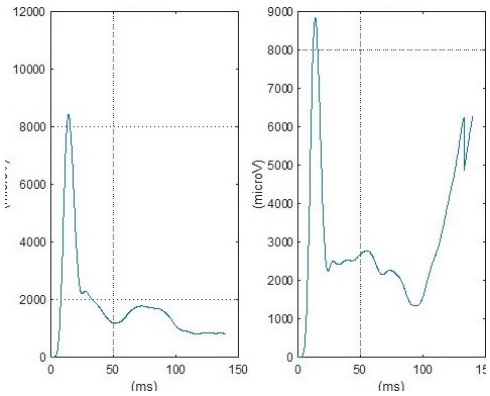


Fig. 4: A signal pair whose morphology is anomalous.

distribution of P for MEP signal pairs is reported in Fig.5, along with the related boxplot. Then, we discard those signal pairs whose P represents an outlier of the sample distribution [25]:

$$T \geq Q_3 + 1.5 \cdot (Q_3 - Q_1) \tag{3}$$

where Q_1 (Q_3) is the 25-th (75-th) percentile of the P sample distribution. For example, in the case of MEP signals, 8% of available signal pairs are discarded. Similar results are obtained on rCMAP data.

3) In order to achieve a straightforward measure sensitive to morphological differences in pre/post signal pairs, we evaluate the energy of error signal:

$$E = |E_{pre} - E_{post}| \tag{4}$$

with E_{post} and E_{pre} being the energy values of pre/post signal pairs. Again, we have worked out the sample distribution of E for peripheral and central signals separately. Then, signal pairs whose error energy represents an outlier of this sample distribution are discarded.

After this pre-processing stage, ITVs are evaluated, and all the steps described in Sect. III-A are performed, achieving a *corrected Global Score*.

C. Contribution 2: Reliability of evoked potential scores

Given that the data obtained from MEP/rCMAP signal analysis are critical, we believe that it is crucial to devise a proper reliability metric. To this end, Dynamic Time Warping (DTW) [26] is employed to assess the reliability of measures based on MEP/rCMAP signals. DTW is a non-linear normalization technique used to quantify similarity between two signals. The algorithm takes two sequences and aligns them so as to minimize their Euclidean distance, and has been profitably addressed to analyze video, audio and graphic data sequences, as well as to compare gait patterns [9], [10]. In this work, it is applied to compare pre/post signal pairs. It returns two outputs: the *DTW distance*, that is Euclidean distance between the two aligned sequences, and the *Warping Index* (WI), the number of sample repetitions needed by the DTW to achieve optimum realignment [10]. Another possible reliability measure is the number of signals pairs that have been discarded due to atypical characteristics. Hence, for each PwMS, and for each pre/post signal pair of both central and peripheral evoked potential measures, we evaluate:

- Average DTW Distance;
- Average Warping Index;
- Number of discarded signal pairs NP.

Large values of these measures denote gross differences in the pre/post waveforms, hardly motivated from the physiopathological point of view. For the sake of simplicity, these measures have been combined in order to achieve a single reliability metric. The sample distribution of each measure has been evaluated over the 100 available samples. Then, DTW distance and WI have been compared with quartile intervals of their sample distribution. An integer score ranging from zero to three is assigned, depending on the distribution quartile interval the value belongs to, and the results are combined together, weighted with NP and normalized so as to achieve a global reliability score REL, ranging between 0 (highly

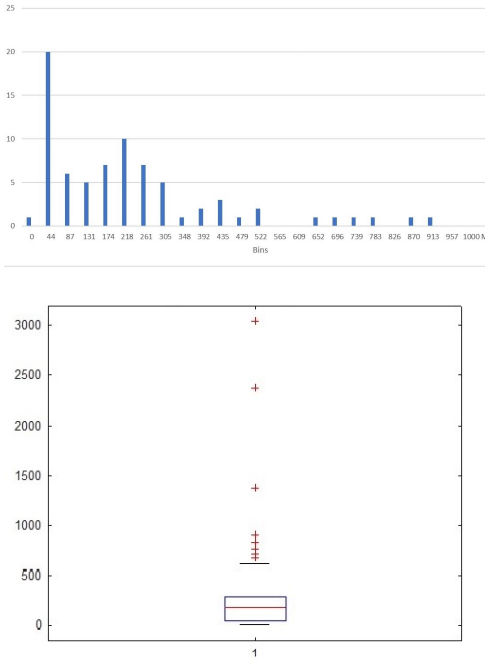


Fig. 5: (a): Sample distribution of peak difference in MEP signal pairs. (b): Related boxplot.

unreliable measure) and 100 (highly reliable measure).

We understand that the sample distributions of DTW parameters may be unstable due to the limited number of available signals; however, the goals of this work is to highlight potential problems related to the use of evoked potentials; hence, this reliability metric helps shedding some light on potential instabilities of such measures.

D. Contribution 3: motor and subjective indices

The information gathered via automatic evaluation of evoked potentials can be compared to gait data and subjective scores. In order to ease such a comparison, for each of these two classes we have worked out proper metrics to be automatically evaluated.

Motor Index, based on walking tests. Data related to the T-25FWT and 6MWT have been measured on PwMS equipped with a smartphone fixed at their waist in position close to the center of mass. As already discussed, even though the gold standard motion capture system [23] was not available for this preliminary study, and considered that our purpose was to gather very simple gait measures to compare with motor evoked potentials and subjective scales, the use of low-cost, easy-to-use instrumentation such as smartphones is reasonable. At present, the gait data measured on the PwMS have only been used to evaluate the time required for T-25FWT and the distance covered in the 6MWT. However, much more information is likely to be embedded in such inertial data (see Sect. II), whose exploitation is left for future developments.

Indices of possible improvement/impairment of the PwMS have been worked out following [14].

- As for the T-25FWT, a score equal to 1 (0, -1) has been assigned if the test performed after drug administration turned out to be improved (stationary, impaired) by a given percentage, i.e. the PwMS walked the 25-feet distance in significantly less time after fampridine administration. In [14], the threshold to decide for an improvement was set to 30%. However, in this paper, and following also [5], we have set a 15% threshold, because 30% turned out to be too demanding and leading to very few patients classified as improved.
- Similarly, a 6MWT score 1 (0, -1) if the PwMS was able to improve significantly the meters walked in 6 minutes after drug administration (15% threshold).
- Finally, the **Motor Index MI** is defined as the average of these two scores (values: ± 1, ± 0.5, 0, positive values denoting improvement). For the sake of clarity, in Table I not only MI, but also the absolute percentage differences in the two walking tests, are reported.

Subjective Index, based on fatigue questionnaires. We have decided to base our evaluation on the MFIS and FSS questionnaires, as not all the PwMS involved in these experiments filled in the VAFS scale. We evaluate the percentage difference in the MFIS and FSS questionnaires before and after fampridine administration. A score equal to 1 (0, -1) has been assigned if the questionnaires filled in after drug administration point out an improved (stationary, impaired) fatigue sensation by a given percentage (set to 15 %). Then, we define the **Subjective Index SI** as the average of these two scores (values: ± 1, ± 0.5, 0, positive values denoting improvement in subjective indices). For the sake of clarity, in Table I not only SI, but also the absolute percentage differences in the two questionnaires, are reported.

IV. Results and discussion

Table I summarizes the obtained results on the ten PwMS addressed in this paper, in terms of:

- GS: Global Score based on evoked potentials, described in Sect. III-A. We recall that this index ranges between -6 and 6, and that, if is exceeds 2, the PwMS is classified as responsive to fampridine.
- cGS: Corrected global score, based on pre-processing and preliminary data selection. Also this index ranges between -6 and 6 and is directly comparable to the previous one.
- REL: Reliability Index (Sect. III-B).
- MI: Motor Index (Sect. III-D).
- SI: Subjective Index (Sect. III-D).
- Absolute percentage motor variation ΔM for either motor test (see Sect. III-B).
- Absolute percentage subjective variation ΔS for either fatigue questionnaire (see Sect. III-D).

We can make the following consideration.

- GS and cGS carry different, yet rather coherent information. Actually, the two metrics yield the same conclusions as for fampridine responsiveness, using the criterion described in Sect. III-B, and they differ at most of ± 1. This denotes that the instability sources in the latency and area evaluation are not sufficient to lead to erroneous conclusions based on the described criteria.
- Nevertheless, the reliability values are quite sparse, ranging from 100% (PwMS no. 10) to 41% (PwMS no. 8). In this latter case, most muscle pairs turned out to be unreliable for motor evoked evaluation. This is a very important result of our work, to be taken into proper account when taking clinical decisions on a specific patient. The clinical use of such information is left to future developments, along with a better statistical assessment involving a larger number of patients and/or different reliability metrics.
- The subjective and motor indices are not always coherent with motor evoked metrics. This may depend on several reasons. As for SI, the questionnaires consider several aspects of fatigue (not only muscular) that may sometimes dominate the overall quantification. In other words, motor conditions of a given patient may be objectively improved by fampridine, but his/her perceived fatigue may not be improved due to non muscular (e.g. psychological) aspects. As for the MI, it must be noticed that the results strictly depend on the EDSS level of the single PwMS. If the EDSS score is very low (e.g. PwMS no. 6), the MI may not suggest improvements because the patient is actually not limited in the walking task. Hence, it is difficult to appreciate a muscular fatigue in this case.

PwMS	EDSS	GS	cGS	REL	MI	SI	ΔM_{25F} (%)	ΔM_{6M} (%)	ΔS_{MFIS} (%)	ΔS_{FSS} (%)
1	5.5	1	1	91	-1	0	-26	-28	1	NA
2	6	0	0	68	0	0	-7	0	+6	+2
3	6	2	3	82	-0.5	NA	-17	(*)	NA	NA
4	3	2	2	65	0	1	+2	+5	+15	+37
5	4.5	2	2	59	-0.5	+0.5	-15	0	+4	+17
6	1	0	1	95	0	0	+4	-2	-2	-9
7	3.5	1	0	91	0	0	0	+12	0	0
8	6.5	0	0	41	1	1	+34	+59	+15	+52
9	3	0	0	73	0	0.5	+3	+13	+17	+11
10	6.5	0	0	100	0	-1	-6	0	-21	- 24

TABLE I: Objective fatigue evaluation metrics. GS (cGS): evoked potentials Global Score (corrected Global Score); range: [-6,6]. REL: Reliability Index; range: [0% - 100%]; MI: Motor Index; range: [-1 1], positive denoting improvement. SI: Subjective Index; range: [-1 1], positive denoting improvement. ΔM_{25F}, ΔM_{6M}: percentage differences in T-25FWT and 6MWT respectively, before and after fampridine administration. ΔS_{MFIS}, ΔS_{FSS}: percentage differences in the MFIS and FSS respectively, before and after fampridine administration. NA: not available. (*): test not finished.

On the other hand, if the disability degree is severe, the patient may not be able to conclude the test and/or the results may be intrinsically unreliable as the high disability level acts as a confounding factor. This suggests that MI is a reliable metric only in intermediate EDSS situations; the validation of this claim is left to future developments.

A. Statistical analysis

Even though the number of treated PwMS is limited, we have worked out the Spearman correlation coefficients ρ between GS/cGS and the motor index MI, as well as SI, after applying Bonferroni correction. GS and cGS turned out to be significatively correlated: $\rho = 0.94$ ($p = 0.0012$, $95\%CI = [0.52, 0.96]$), as well as GS/cGS and MI ($\rho = 0.84$, $p = 0.003$, $95\%CI = [0.56, 0.91]$). On the other hand, the correlation between GS/cGS and SI cannot be assessed as it does not reach significativity ($p = 0.2$).

V. Conclusions

In this paper, we have exploited data gathered at the Regional Multiple Sclerosis Reference Centre, 2nd Neurology Department of San Luigi Gonzaga Hospital, Orbassano, Italy, within a clinical trial to assess the responsiveness to fampridine for fatigue relief. Ten PwMS have been subject to transcranial magnetic stimulation and high voltage electrical stimulation to record MEPs and rCMAPs on ten leg muscle locations. Moreover, participants performed the T-25FWT and 6MWT. Then, fampridine was administered for two weeks, and all the tests repeated. The related signals have been manually evaluated by clinicians.

We have considered information from central and peripheral motor evoked potentials, as well as results of walking tests and subjective questionnaires filled in by PwMS. We have worked out an algorithm to automatically evaluate latency/area ITVs, limiting some identified causes of instability. We have proposed a measure of reliability of such metrics, evaluated by means of the DTW algorithm. We have proposed a Motor Index and a Subjective Index, which concisely represent the information related to walking tests and subjective questionnaires respectively.

The significance of this work lies in the intrinsic importance of a steady, objective fatigue evaluation, with potential impact in the assessment of therapy efficacy, hence in the quality of life of PwMS.

The most important obtained result concerns the reliability of evoked potential measures. Even though the global index proposed in [14] and our corrected index yield rather coherent information, the reliability of such measures is quite sparse, suggesting moderate repeatability of these measures. This enforces the belief that MEP/rCMAP potentials as a technique to assess fatigue are intrinsically unsteady even though a proper protocol is applied to limit confounding factors.

Due to the limited number of PwMS involved in this stage of the work, these results should be interpreted as preliminary indications that could drive subsequent analysis. Nevertheless, we have achieved the main goals of this preliminary work, namely: to assess motor potential reliability, and to investigate whether different metrics yield coherent information related to fatigue.

As future developments, we will recruit more PwMS in the trial, so as to assess the methods form the statistical point of view. We will try to get more information from the walking tests, e.g. related to postural stability, useful not only for the fatigue evaluation but also for the follow up of PwMS.

References

[1] C. H. Polman, S. C. Reingold, B. Banwell, M. Clanet, J. A. Cohen, M. Filippi, K. Fujihara, E. Havrdova, M. Hutchinson, L. Kappos, F. D. Lublin, X. Montalban, P. l O'Connor, M. SandbergWollheim, J. Thompson, E. Waubant, B. Weinshenker, and J. S. Wolinsky, "Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria," *Annals of Neurology*, vol. 69, pp. 292302, 2011.

[2] E. Leray, T. Moreau, A. Fromont, and G. Edan, "Epidemiology of multiple sclerosis," *Elsevier Masson Revue Neurologique*, vol. 172, pp. 3–13, 2015.

[3] A.V., "Fatigue and multiple sclerosis: evidence-based management strategies for fatigue in multiple sclerosis," *Multiple sclerosis council for clinical practice guidelines*, pp. 1 – 8, 1998.

[4] L. Filli, J. Werner, G. Beyer, K. Reuter, J. A. Petersen, M. Weller, B. Zorner, and M. Linnebank, "Predicting responsiveness to fampridine in gaitimpaired patients with multiple sclerosis," *European Journal of Neurology*, 2018.

[5] E. Allart, A. Benoit, A.Blanchard-Dauphin, V. Tiffreau, A. Thevenon, H. Zephir, O. Outteryck, A. Lacour, and P. Vermersch, "Sustained-released fampridine in multiple sclerosis: effects on gait parameters, arm function, fatigue, and quality of life," *Journal of Neurology*, vol. 262, no. 8, pp. 1936–1945, 2015.

[6] P. Flachenecker, T. Kumpfel, B. Kallmann, M. Gottschalk, O. Grauer, P. Rieckmann, C. Trenkwalder, and K. V. Toyka, "Fatigue in multiple sclerosis: a comparison of different rating scales and correlation to clinical parameters ," *Multiple Sclerosis Journal*, pp. 1 – 8, 2002.

[7] M. T. V. Chan, "Current status of motor evoked potentials," *JNACC Conference proceedings*, vol. 3, no. 4, pp. 59–61, 2016.

[8] S. H. Salzman, "The 6-Min Walk Test," *Chest Journal*, vol. 135, no. 5, pp. 13451352, 2009.

[9] M. M. Engelhard, S. R. Dandu, , S. D. Patek, J. C. Lach, and M. D. Goldman, "Quantifying six-minute Walk Induced Gait Deterioration with Inertial Sensors in Multiple Sclerosis Subjects," *Elsevier Gait and Posture*, vol. 49, pp. 340–345, 2016.

[10] S. R. Dandu, M. M. Engelhard, A. Qureschi, J. Gong, J. C. Lach, M. Brandt-Pearce, and M. D. Goldman, "Understanding the Physiological Significance of Four Inertial Gait features in Multiple Sclerosis," *Journal of Biomedical and Health Informatics*, , no. Special issue, pp. 2168–2194, 2017.

[11] W. Troni, A. Di Sapio, E. Berra, S. Duca, A. Merola, and F. Sperli, "A methodological reappraisal of non-invasive high voltage electrical stimulation of lumbosacral nerve roots," *Clinical Neurophysiology*, pp. 71–80, 2011.

[12] A. Di Sapio, A. Bertolotto, F. Melillo, F. Sperli, S. Malucchi, and W. Troni, "A new neurophysiological approach to assess central motor conduction damage to proximal and distal muscles of lower limbs," *Clinical Neurophysiology*, vol. 125, pp. 133–141, 2014.

[13] W. Troni, F. Melillo, A. Bertolotto, S. Malucchi, M. Capobianco, and F. Sperli, " Normative Values for Intertrial Variability of Motor responses to Nerve Root and Transcranial Stimulation: A Condition for Follow-Up Studies in Individual Subjects," *PLoS ONE*, vol. 11, no. 5, pp. 1–18, 2016.

[14] A. Di Sapio, S. Giorgi, A. Bertolotto, F. Sperli, F. Melillo, and W. Troni, "Effects of fampridine administration on central motor conduction failure in multiple sclerosis patients," *Elsevier Clinical Neurophysiology*, vol. 129, pp. 18, 2018.

[15] S. Samar, M. Ayacheabc, and A.Chalahab, "Fatigue in multiple sclerosis - Insights into evaluation and management ," *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 47, no. 2, pp. 139–171, 2017.

[16] J. Chilcot, S. Norton, M. E. Kelly, and R. Moss-Morris, "The Chalder Fatigue Questionnaire is a valid and reliable of perceived fatigue severity in multiple sclerosis," *Multiple sclerosis journal*, pp. 1 – 8, 2017.

[17] M. Psarakis1, D. Greene, M. H. Cole, S. R. Lord, P. Hoang, and M. Brodie, " Wearable technology reveals gait compensations, unstable walking patterns and fatigue in people with multiple sclerosis," *Institute of Physics and Engineering in Medicine Physiological Measurement*, vol. 39, no. 7, 2018.

[18] A. Kalronab, "The correlation between symptomatic fatigue to definite measures of gait in people with multiple sclerosis ," *Gait and Posture*, vol. 44, pp. 178–183, 2016.

[19] R. Sacco, R. Bussman, P. Oesch, J. Kesselring, and S. Beer, "Assessment of gait parameters and fatigue in MS patients during inpatient rehabilitation: a pilot trial," *J. Neurol.*, vol. 258, pp. 889–894, 2011.

[20] Y. Terao, Y. Ugawa, K. Sakai, Y. Uesaka, N. Kohara, and I. Kanazawa, "Transcranial stimulation of the leg area of the motor cortex in humans," *Acta Neurol Scand*, pp. 78–83, 1994.

[21] R. A. Marrie M. D. Goldman and and J. A. Cohen, "Evaluation of the six-minute walk in multiople sclerosis subjects and healthy controls," *Multiple Sclerosis*, vol. 14, pp. 383–390, 2008.

[22] L. A. Nogueira, L. Teixeira, P. Sabino, H.A. Filho, R.M. Alvarenga, and L.C. Thuler, "Gait characteristics of multiple sclerosis patients in the absence of clinical disability," *Disabil Rehabil*, vol. 35, pp. 472–1478, 2013.

[23] P. Merriaux, Y. Dupuis, R. Boutteau, V. Vasseur, and X. Savatier, "A Study of Vicon System Positioning Performance ," *Sensors*, vol. 17, pp. 1–18, 2017.

[24] P. Sao, R. Hegadi, and S. Karmakar, "A Literature Review on Approaches of ECG Pattern Recognition," *Int. Journal of Information Science and Intelligent System*, vol. 3, pp. 79–90, 2014.

[25] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*, Springer Verlag, 2013.

[26] Q. Li and G. D. Clifford, "Dynamic time warping and machine learning for signal quality assessment of pulsatile signals," *Physiological Measurement*, vol. 33, pp. 1491 –1501, 2012.