Comparing recalibration strategies for electroencephalography-based decoders of movement intention in neurological patients with motor disability

Eduardo López-Larraz[1,2,*], Jaime Ibáñez[3,4,*], Fernando Trincado-Alonso[5], Esther Monge-Pereira[6], José Luis Pons[4,7], Luis Montesano[2,8]

[1]*Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, Silcherstr. 5, 72075, Tübingen, Germany (Correspondence address).*
[2]*Instituto de Investigación de Ingeniería de Aragón (I3A) and Departamento de Informática e Ingeniería de Sistemas, University of Zaragoza, Zaragoza, Spain.*
[3]*Sobell Department of Motor Neuroscience, Institute of Neurology, University College London, London, UK.*
[4]*Neural Rehabilitation Group, Cajal Institute, Spanish National Research Council (CSIC), Madrid, Spain.*
[5]*Biomechanics and Technical Aids Unit, Hospital Nacional de Parapléjicos, Toledo, Spain.*
[6]*Departamento de Fisioterapia, Terapia Ocupacional, Rehabilitación y Medicina Física, Universidad Rey Juan Carlos, Alcorcón, Spain.*
[7]*Tecnológico de Monterrey, Mexico.*
[8]*Bit&Brain Technologies SL, Zaragoza, Spain.*
[*]*Equal contributors.*
Correspondence e-mail: eduardo.lopez-larraz@uni-tuebingen.de

Motor rehabilitation based on the association of electroencephalographic (EEG) activity and proprioceptive feedback has been demonstrated as a feasible therapy for patients with paralysis. To promote long-lasting motor recovery, these interventions have to be carried out across several weeks or even months. The success of these therapies partly relies on the performance of the system decoding movement intentions, which normally has to be recalibrated to deal with the non-stationarities of the cortical activity. Minimizing the recalibration times is important to reduce the setup preparation and maximize the effective therapy time. To date, a systematic analysis of the effect of recalibration strategies in EEG-driven interfaces for motor rehabilitation has not yet been performed. Data from patients with stroke (4 patients, 8 sessions) and spinal cord injury (4 patients, 5 sessions) undergoing two different paradigms (self-paced and cue-guided, respectively) are used to study the performance of the EEG-based classification of motor intentions. Four calibration schemes are compared, considering different combinations of training datasets from previous and/or the validated session. The results show significant differences in classifier performances in terms of the true and false positives. Combining training data from previous sessions with data from the validation session provides the best compromise between the amount of data needed for calibration and the classifier performance. With this scheme, the average true (false) positive rates obtained are 85.3% (17.3%) and 72.9% (30.3%) for the self-paced and the cue-guided protocols, respectively. These results suggest that the use of optimal recalibration schemes for EEG-based classifiers of motor intentions results in enhanced performances of these technologies, while not requiring long calibration phases prior to starting the intervention.

Keywords: Brain-machine interfaces (BMI); Movement intention; Rehabilitation; Stroke; Spinal cord injury (SCI).

1. Introduction

The association of mental states related to motor actions with proprioceptive coherent feedback constitutes a feasible strategy to boost functional rehabilitation of patients with neurological injuries such as stroke or spinal cord injury (SCI).[1] In fact, it has been proven as the only means to facilitate recovery for certain pathological conditions such as in chronic stroke patients with no residual movement.[2,3] Technologies based on electroencephalographic (EEG) signals are of special interest, since they allow the precise and real-time estimation of cortical activation over motor areas,[4–6] and at the same time, can be used in clinical environments. A large number of studies have been carried out during the past years proving the benefits of using interventional strategies based on brain-machine interface (BMI) technologies relying on the EEG activity.[3,7–12] These experimental protocols involve multiple BMI sessions along several weeks to induce significant plastic changes (e.g., cortico-cortical and cortico-muscular connectivity) and promote functional recovery (e.g., Fugl-Meyer index, American spinal injury association impairment scale) in the patients. The encouraging results observed so far in terms of functional and neurophysiological improvements in patients undergoing BMI interventions[3,11,12] support the need for further validating these technologies in clinical environments, and reducing the associated limitations of current solutions.

Two typically critical aspects in BMIs for rehabilitation are the time needed to set up the technology on the patient and the impact on the systems' performance of non-stationarities in EEG signals within single sessions and, more importantly, across different recording days.[13] If BMI-based interventions are to be applied on patients along multiple sessions, a major goal will be to maximize the effective BMI-driven rehabilitation time. From a practical point of view, for a patient that attends to a one-hour BMI rehabilitation session, the idea would be to maximize the time in which the patient receives associative feedback. This can be related to: *i)* the time needed to calibrate the EEG decoding systems at the beginning of each training session, before the patients start the actual BMI intervention (the less time needed for this part, the better); and *ii)* the percentage of intervention trials with correct estima-

tions of the movement-related cortical activity, which lead to a coherent proprioceptive feedback (a strong correlation has been reported between the subject-specific accuracy of the EEG classifiers of motor intentions and the induced plastic changes after the BMI intervention[14]). In this regard, the calibration of signal processing and classification algorithms to decode movement-related mental states is a critical element to consider.

There are two major strategies to detect motor intentions from EEG signals. On the one hand, these can be decoded from the event-related desynchronization (ERD) of sensorimotor rhythms, which consists of a decrease in the energy of cortical rhythmic activity within the mu (8-12 Hz) and beta (13-30 Hz) frequency bands, starting around 1.5 s before planned motor actions.[15] On the other hand, movement intentions are also reflected in slow (below 1 Hz) variations of the EEG amplitudes around the cranial vertex, which are referred to as movement-related cortical potentials (MRCP). MRCPs in planned movements start with a negative deflection of the EEG amplitude at approximately the onset of the movements, known as the Bereitschaftspotential (BP).[16] Both ERD and MRCP cortical patterns have been used for BMI applications with neurological patients, since they are reliable sources of information to estimate in real-time when a subject is about to start a movement.[17–21] Recalibration of systems using either of these patterns require considering changes in the EEG signal properties (signal-to-noise ratios and non linear dynamics of the signal principally) and in the state of the subjects (changes in concentration, fatigue, drowsiness, etc), which can influence the efficiency of the algorithms to detect motor intentions.[13]

Previous studies have dealt with the recalibration of BMI systems to provide subjects with alternative communication channels.[13] A study by Niazi et al. proposed a way to design universal decoders of motor intentions for neurorehabilitation purposes, but classification results obtained with patients decreased substantially compared to healthy subjects.[22] This was likely due to the high heterogeneity in cortical activation patterns typically observed in patients with damages in the central nervous system.[23–25] To date, only very preliminary works have explored the influence of recalibration strategies for BMIs detecting motor intentions with

neurological patients, pointing the importance of using recent data to improve the performance.[26–28]

This study evaluates four different recalibration strategies for BMI systems detecting motor intentions based on the identification of ERD and MRCP patterns. All these strategies are feasible solutions in multi-session BMI interventions carried out with patients in clinical facilities. To reach generalizable results that can be directly extrapolated to any BMI platform dealing with motor intention detection for neurological rehabilitation, two datasets from experiments carried out with patients are analyzed offline. The first dataset is from a group of chronic stroke patients who participated in an 8-sessions BMI intervention. The second dataset is from a group of subacute patients with SCI who underwent 5 BMI sessions. The studied recalibration strategies rely on the use of data from the past and/or the validated session to calibrate the movement intention decoder.

2. Methods

The datasets used to compare the different BMI recalibration strategies were acquired in two sets of experiments for arm/hand rehabilitation in patients with chronic stroke (Exp1) and subacute SCI (Exp2). Table 1 summarizes the details of the two datasets, which are described in sections 2.1 and 2.2. In both experiments, patients underwent a multiple-sessions intervention based on detecting motor intentions from EEG signals to provide a coherent (visual and/or proprioceptive) feedback. Two different BMI paradigms were used: stroke patients (Exp1) were asked to perform self-initiated movements and SCI patients (Exp2) followed a cue-guided protocol. As commonly done in BMI research applied to neurorehabilitation, patients first performed several repetitions of the desired task (i.e., attempting to move their paralyzed limb) without feedback to calibrate the BMI in each session. Then, they performed some closed-loop trials in which they were given feedback (functional electrical stimulation to assist the movements trained) according to their brain activity. This study refers only to the data recorded during the calibration trials, since it allowed us to perform offline comparisons with data not affected by the effects of the stimulation. Trials with feedback were not eligible to compare calibration schemes; firstly, to avoid possible influences of the afferent stimulation on the EEG signals; and secondly, because they were influenced by the type of calibration scheme used to train the BMI triggering the proprioceptive stimuli.

Both experimental protocols considered here were designed in accordance with the Declaration of Helsinki, and approved by the corresponding Local Ethics Committees (Universidad Rey Juan Carlos, Madrid, Spain, for Exp1; and Hospital Complex of Toledo, Toledo, Spain, for Exp2). All patients signed a written informed consent before their participation in the study.

Table 1. Datasets description

Dataset	Number of patients	Number of sessions	Exp. paradigm	Mov. onset detection	Trials recorded per session
Exp1	4 Stroke	8	Self-paced	Gyroscopic data	35
Exp2	4 SCI	5	Cue-guided	Visual cues	40

2.1. *Dataset for Exp1*

2.1.1. *Patients*

Four chronic stroke patients (all males, age 54 ± 12 years, mean $\pm$ SD) with a lesion in the territory of the middle cerebral artery and a predominance of brachial hemiparesis were recruited (see Table 2). The experimental sessions were carried out at the Universidad Rey Juan Carlos, Madrid (Spain).

Table 2. Details of stroke patients

ID	Age (years)	Time since lesion (years)	Type of stroke	Gender	Lesioned hemisphere	Fügl-Meyer	Rh sessions a week
P1	69	4	Hemorr.	Male	Right hem.	64	0
P2	54	4	Hemorr.	Male	Right hem.	69	2
P3	54	3	Ischem.	Male	Left hem.	68	2
P4	40	13	Hemorr.	Male	Right hem.	81	2

2.1.2. *Data acquisition*

EEG was recorded from 32 electrodes placed at AFz, F3, F1, Fz, F2, F4, FC5, FC3, FC1, FCz, FC2, FC4, FC6, C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4, CP6, P3, P1, Pz and P4 (according to the international 10/10 system). The ground and reference electrodes were placed on FPz and on the left earlobe, respectively. Active Ag/AgCl electrodes (actiChamp, Brain Products GmbH, Germany) were used. EEG was amplified and digitized using a g.Tec amplifier (Guger Technologies, Graz, Austria) at a sampling rate of 256 Hz, and power-line notch-filtered at 50 Hz.

The arm and hand movements were measured with solid-state gyroscopes and surface electromyo-

graphy (EMG). Three gyroscopes (Technaid S.L., Madrid, Spain), placed on the hand dorsum, the distal third of the forearm, and the middle of the arm measured the limb kinematics. The data were sampled at 100 Hz (gyroscopes) and 1000 Hz (EMG). A common digital signal was acquired by all systems and posterior offline resampling was carried out to store all measured signals synchronized at 256 Hz.

2.1.3.   *Experimental design*

The experimental protocol consisted of eight sessions performed during one month (two sessions per week). In each session, the patients first performed (or attempted) 35 reaching movements with the affected arm. Patients were instructed to perform self-paced reaching movements separated by resting intervals of at least 5 seconds. The resting phases between movements ensured that each new trial was initiated from a baseline condition and that there was enough time provided to record the entire ERD and BP patterns (which are supposed to start around 1.5 s before the movement onset). The movement intervals had a duration between 4 and 5 s, while the resting intervals were around 6-10 s long. EEG data of these movements was used to train an EEG-based system that based the estimations of movement intentions on the combined characterization of the ERD and the MRCP cortical patterns. During the intervention phase (data not considered in this study), patients performed/attempted 80 movements, and electrical stimuli were delivered each time movement intentions were detected from the EEG data.

2.1.4.   *Definition of movement events*

Since the experimental protocol used for Exp1 was self-paced (the patients initiated the movements at their own chosen pace), the localization of movement events was required to train and evaluate the BMI. The onsets of the performed movements were located by processing the gyroscopic data, which were more reliable than EMG signals and, in addition, have proven to be reliable to precisely locate the BP peak.[6] To detect the time at which each movement started, the gyroscopic sensor at the hand dorsum was analyzed. Data were low-pass filtered (Butterworth, order 2, < 6 Hz), and the peak amplitude was estimated for each subject performing the reaching movement. The threshold amplitude for the detec-

tion of the onsets of the movements was set to 5% of this peak amplitude and posterior visual correction was performed to assure that the events were correctly labeled.

2.2.   *Data description for Exp2*

2.2.1.   *Patients*

Four patients with incomplete tetraplegia participated in this experiment. All of them were in a subacute stage, and were hospitalized at the Hospital Nacional de Parapléjicos, in Toledo (Spain), where the experimental sessions took place. Clinical details of each patient can be seen on Table 3.

Table 3.   Details of SCI patients

ID	Age (years)	Time since lesion (months)	Type of lesion	Gender	Dominant hand	Stimulated hand
P1	71	4	C5, ASIA C	Male	Right	Left
P2	38	9.5	C5, ASIA C	Male	Right	Left
P3	36	7	C5, ASIA B	Male	Right	Right
P4	55	4	C4, ASIA C	Male	Right	Right

2.2.2.   *Data acquisition*

EEG was recorded from 32 electrodes placed at AFz, F3, F1, Fz, F2, F4, FC5, FC3, FC1, FCz, FC2, FC4, FC6, C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4, CP6, P3, P1, Pz and P4 (according to the international 10/10 system). The ground and reference electrodes were placed on FPz and on the left earlobe, respectively. Active Ag/AgCl electrodes (g.LADYbird, Guger Technologies, Graz, Austria) were used. EEG was amplified and digitized using a g.Tec amplifier (Guger Technologies, Graz, Austria) at a sampling rate of 256 Hz, and power-line notch-filtered at 50 Hz.

2.2.3.   *Experimental design*

The experimental protocol consisted of five sessions, performed within a maximum time of 10 days. In each session, the patients performed 40 trials in which they were asked to try to close one of their hands (see rightmost column in Table 3) without any feedback. They were executed in 20-trial blocks, with visual cues indicating to the patient what to do on each moment. The trials started with a "Rest" cue (i.e., the word Rest with random duration between 4 and 7 seconds), followed by a "Movement" cue (i.e.,

the word Movement, during 3 seconds). These visual cues were displayed in a screen placed 1–1.5 m away from the patient. Subsequently, they performed between 40 and 80 closed-loop trials in which they were provided with electrical and visual feedback (i.e., electrical stimulation of the finger flexors to close the hand and a virtual hand in the screen that closed when the stimulation was delivered) when the intention of movement was detected by the BMI (data not considered in this study).
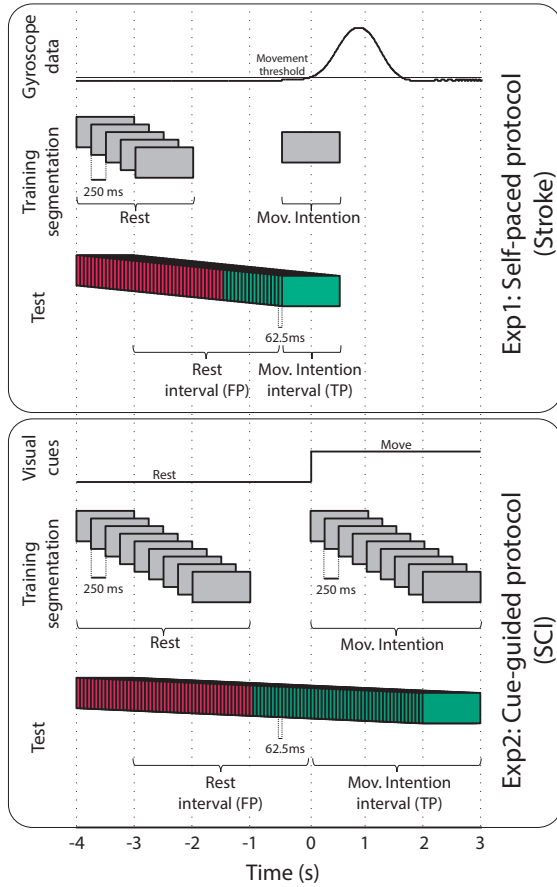

Figure 1. Application of the proposed EEG-based classifier of motor intentions on the data recorded with the asynchronous (top) and the cue-guided (bottom) paradigms. On each panel, the first row corresponds to the data used to define the movement onset; the second row shows the one-second windows used to train the classifier with examples of the *Rest* and *Mov. Intention* classes; and the third row represents the one-second sliding windows used for testing, with a sliding step of 62.5 ms. In this third row, the time intervals corresponding to Rest ([-3, -0.5] for Exp1 and [-3, 0] for Exp2) and Mov. Intention ([-0.5, 0.5] for Exp 1 and [0, 3] for Exp2) are also marked. The test windows are colored in red if they correspond to Rest and in green if they correspond to

Mov. Intention. Notice that the color of the window is based on its ending time: *i.e.*, the window covering the interval [-0.9, 0.1] would be colored in green, as it actually corresponds to the Mov. Intention period.

2.3. EEG-based classifier of motor intentions

Despite the two datasets studied in this paper correspond to different paradigms (self-paced and cue-guided), the decoder of motion intention is based on the same principles in both cases, and is trained with examples of *rest* and *movement intention* signals in order to predict new unseen data (Fig. 1).

2.3.1. Data preprocessing

The EEG signals were first trimmed down to 4.5-second trials for Exp1 (i.e., from -4 to 0.5 s, with respect to the movement onset), and to 7-second trials for Exp2 (i.e., from -4 to 3 s, with respect to presentation of the cue). Trials containing artifacts were discarded after visual inspection.

2.3.2. Feature extraction

The BMI decoder was designed to distinguish between *rest* and *movement intention* brain states. A one-second long sliding window was used to simulate an online operation of the BMI. Since the experiments relied on different paradigms (i.e., self-paced for Exp1 and cue-guided for Exp2), we used different time windows to calibrate the decoder. For the dataset of Exp1, the *rest* interval was defined as [-4, -2] s and the *movement intention* interval was defined as [-0.5, 0.5] s, with $t = 0$ being the onset of the movement (see Fig. 1, top). This strategy was followed to maximize temporal precision of the motor intention detections, which is one of the most desired properties for self-paced BMI protocols.[14] For the dataset of Exp2, the *rest* interval was defined as [-4, -1] s and the *movement intention* interval was defined as [0, 3] s, with $t = 0$ being the presentation of the "Movement" cue (see Fig. 1, bottom). In this case, the objective was to maximize the number of trials with the movement intention correctly decoded during the "Movement" period. Within these intervals, the one-second window was applied with a sliding step of 0.25 s to obtain training examples of *rest* and *movement intention* classes, and to extract the features.

The decoder was based on both the ERD and the MRCP cortical patterns. The features were extracted from the 21 electrodes shared by both montages (i.e., F3, Fz, F4, FC3, FCz, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P3, Pz, P4). An automatic feature selection system was used to obtain the most discriminant ones. ERD features were computed after applying a Laplacian filter to the signals, and using a 16th order autoregressive model to obtain the power values in the frequency range [7-30] Hz using 1-Hz bins (i.e., resulting in 24 features per channel and window). The MRCP features were extracted after applying a common average reference, bandpass filtering the signals to [0.1-1] Hz and downsampling them to 64 Hz (i.e., 64 features per channel and window).

As a result of this process, a total of 1848 features were extracted for each one-second window, combining power values (i.e., ERD features) and signal amplitudes (i.e., MRCP features). Their values were normalized to have zero mean and unit variance. Sparse discriminant analysis (SDA)[29, 30] was used to select the most discriminative features, removing redundant information, and as linear classifier. The maximum number of features was set to 30, based on previous empirical results.[5] SDA is especially useful for problems with the number of features being large with respect to the number of observations. Using ℓ_1 and ℓ_2 penalties, this method performs a sparsity-based regularization, which allows to significantly reduce the number of features, selecting the most discriminative ones while removing the redundant information. This technique has successfully been used to classify motor commands both in healthy subjects, and in SCI and stroke patients.[5, 20, 31]

2.3.3. *Calibration schemes*

Four different schemes for BMI calibration were implemented and compared for each subject separately. For each scheme, the EEG-based decoder was calibrated using different configurations of datasets, and tested on the data recorded on each session. Each scheme simulates a scenario in which a patient comes on day n for his/her BMI therapy, and the decoder is trained using data obtained on previous and/or the same day.

A) **Previous Session (PrevSes)**: This scheme simulates an intervention in which data for BMI training are recorded on a previous day before starting the therapy. For each session n tested ($n \geq 2$), the BMI is trained with data from session n-1 only.

B) **Accumulated previous sessions (AccumPrevSes)**: This scheme extends the previous one by pooling all the previous sessions together, simulating a database of trials recorded in the past to train the BMI before starting each session. For each session n tested ($n \geq 2$), the training dataset consists of all the sessions previous to the current one (i.e., sessions [1 .. n-1]) pooled together.

C) **Current session (CurrentSes)**: This scheme simulates a scenario in which, on each session, the subjects perform some trials to calibrate the BMI, and then they start the therapy. In this case, data from each session n ($n \geq 2$) are used for training and testing[a], using a leave-one-out cross-validation procedure.

D) **Accumulated sessions (Accum)**: In this scheme, a database of trials from all previous sessions is stored and updated with new data recorded on each therapy session. For each session n ($n \geq 2$), a similar leave-one-out procedure to scheme C is used, but in this case, data from all the previous sessions (i.e., sessions [1 .. n-1]) is appended to the training set.

The comparisons between these four schemes is expected to allow assessing the benefits of using training datasets with: *i*) larger number of samples (*AccumPrevSes* and *Accum* schemes), and *ii*) information obtained from the test/intervention session (*CurrentSes* and *Accum*).

For the two schemes that included training data from the tested session (i.e., *CurrentSes* and *Accum*), we also studied the influence of the amount of data of such session included in the training set. Hence, we computed the performances of both schemes in 5 different configurations, considering only the first 20%,

---

[a]Given that schemes A and B required data from at least one previous session, they were evaluated after session 2. Hence, to facilitate comparisons, schemes C and D were also tested after the second session only.

40%, 60%, 80%, or 100% of the trials. With these reduced datasets, we repeated the same procedure as for the original one: leave-one-out for *CurrentSes*, and modified leave-one-out (including data from all previous sessions in the training set) for *Accum*.

2.3.4. *Optimization of the classifier threshold*

On the test data, a threshold was applied to the output of the decoder to decide at each moment if it corresponded to *rest* or to *movement intention*. This threshold was optimally obtained from the training dataset in an inner 10-fold cross validation loop, following the criterion of maximizing the percentage of trials with a true positive (TP) and with no false positives (FP).

2.3.5. *Decoding evaluation*

All the schemes were tested offline, but simulating an online operation of the system.[32] In each case, the BMI decoder was trained with the corresponding datasets, and the test trials of each session were evaluated with a one-second sliding window with a sliding step of 62.5 ms (i.e., 16 samples at 256 Hz). The decoding performance of each scheme was evaluated by computing the percentage of trials containing true positives (TP) and false positives (FP). TP were defined as the movement intervals (i.e., interval [-0.5, 0.5] for Exp1, and interval [0, 3] for Exp2) with at least one correct *movement intention* output generated by the decoder. FP were defined as the rest intervals (i.e., interval [-4, -0.5] for Exp1, and interval [-4, 0] for Exp2) containing at least one *movement intention* output. Bottom rows on both panels of Figure 1 show how the test trials were evaluated, indicating the rest and movement intervals, and coloring in red and green the windows that belong to each of these intervals, respectively.

In addition, we computed the decoding latency of the TPs as the time of the first *movement intention* output with respect to $t = 0$ (i.e., the onset of the movement for Exp1 and the presentation of the cue for Exp2). This metric can be positive or negative for Exp1 (as, in a self-initiated paradigm, the movement can be decoded before it starts), but only positive for Exp2. The results on detection latencies are provided to ensure that the different recalibration schemes do not modify the detection profiles and to provide an indication of the distribution of the TPs.

2.4. *Statistical analysis*

In order to gain statistical power, results with all patients (stroke and SCI) were used together for the statistical analysis. The analysis was performed on the dependent variables TP, FP, average latencies and standard deviation of latencies. The results with the four calibration schemes were compared to test the hypothesis that they led to significantly different values. Given that the BMI performances with the schemes depended on each subject, a Friedman's test was used. Paired post-hoc comparisons (using the Nemenyi test) were used to analyze significant differences between pairs of calibration schemes. Therefore, it was assumed that the differences in the two experimental paradigms used to obtain the two datasets (stroke and SCI) only have an influence on the absolute performance of the classifier, but not on the relative changes of this performance due to the selection of different calibration schemes. All results are reported as the mean ± SD.

3. Results

On average, the number of available trials on each session after removing artifacts was 28.34±4.80 and 32.25±4.61 for stroke and SCI patients, respectively. Both patient groups presented contralateral cortical activations, including ERD and MRCP, as described in the literature[15, 16] (see Figure 2). The automatic feature selection method primed the use of features in the lower-alpha ([7-8] Hz), upper-alpha ([10-14] Hz) and beta ([18-24] Hz) frequency ranges from both hemispheres, although with a trend towards selecting more contralateral features. This result is in line with previous studies using SDA feature selection.[5, 20] The average features selected for each recalibration scheme for the stroke and the SCI datasets are shown in Figure 3.

In order to evaluate the overall performance of the implemented classifier, the first four rows in Table 4 summarize the average classification results with each recalibration scheme in terms of TP, FP and detection latencies for each of the two paradigms considered. With the *Accum* scheme (which provided the highest average TP/FP ratio), 85.3% and 72.9% of the movements were correctly detected in both groups of patients with average detection latencies of 0.01±0.21 ms (stroke) and 1.22±0.57 ms (SCI). With this scheme, the FP rates were 17.3% and 30.3% for
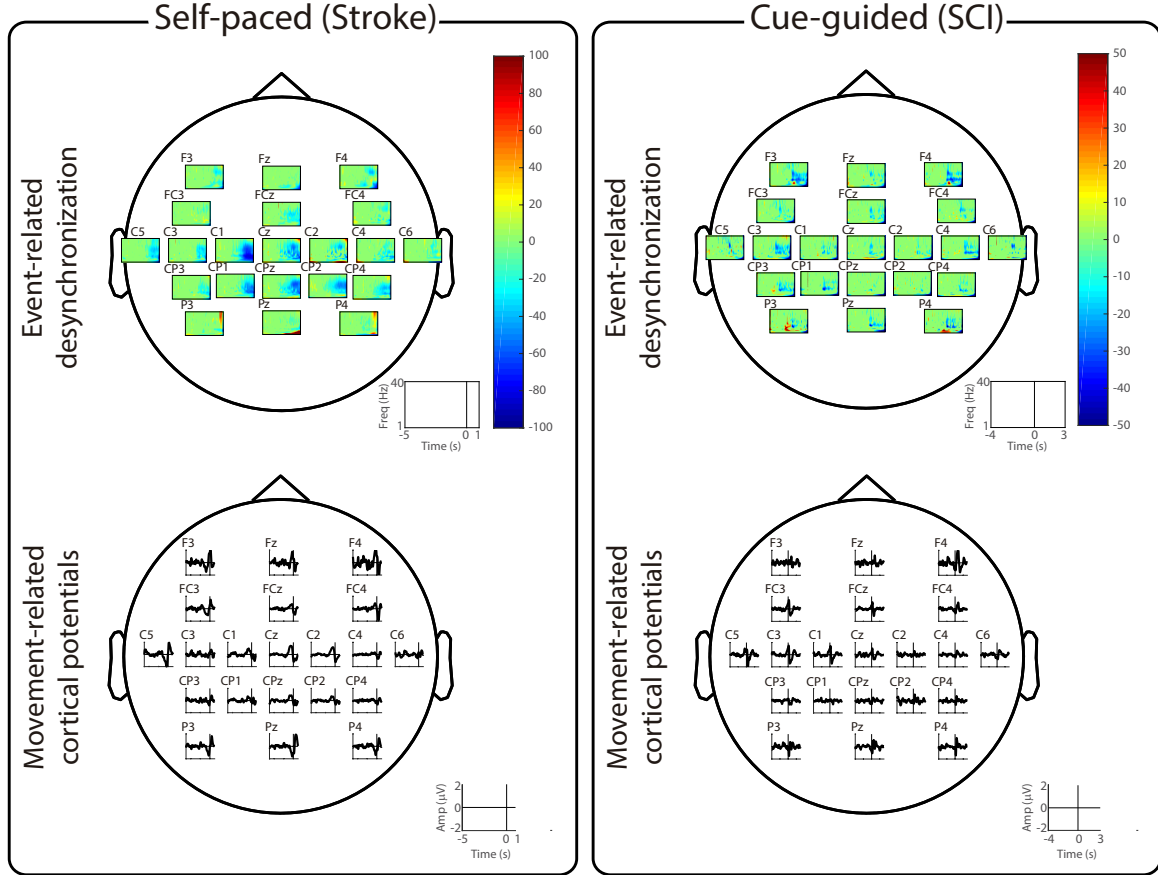
Figure 2.   Cortical activations of both groups of patients. The left panel corresponds to the self-initiated protocol with stroke patients, while the right panel corresponds to the cue-guided protocol with SCI patients. The upper part of each panel represents the event-related desynchronization activity. Notice that the scales are different for both patient groups, since the stroke patients presented an ERD of higher magnitude. The bottom part of each panel represents the movement-related cortical potentials. Note that, for both patient groups, the left- and right-sided electrodes were swapped in the patients that performed the task with the left hand. Therefore, these cortical activation plots represent that, for all the patients, the right arm was the paralyzed one (which was used to perform the movement attempt).

stroke and SCI data, respectively. The fifth row in Table 4 reports the average results of a random classifier. The random classifier was evaluated four times on each dataset (i.e., one for each of the four recalibration schemes), and trained after shuffling the labels of the training trials. This classifier led to notably worse classification results (in terms of TPs and FPs) than those obtained with the four compared schemes using the correct labels.

Figure 4 summarizes the results obtained in each population for each recalibration strategy. In terms of TP, significant differences ($p < 0.001$) were found between the recalibration schemes. Post-hoc paired comparisons showed that the *CurrentSes* scheme led to a significantly higher TP rate than the *Pre-*

vSes and *AccumPrevSes* schemes ($p = 0.003$ and $p = 0.010$, respectively), and that the *Accum* scheme led to a significantly higher TP rate than the *PrevSes* and *AccumPrevSes* schemes ($p = 0.010$ and $p = 0.034$, respectively). Significant differences were also obtained for FP ($p = 0.009$). In this case, the post-hoc analysis showed that the *Accum* scheme led to significantly lower FPs than the *PrevSes* scheme ($p = 0.019$).

Regarding detection latencies, their absolute value for the self-paced protocol was below 50 ms (range -0.07±0.21 – 0.02±0.22 s), while for the cue-guided protocol was above 1 s (range 1.10±0.60 – 1.28±0.58 s). Note that, unlike in the cue-guided paradigms, the self-paced protocol was optimized to
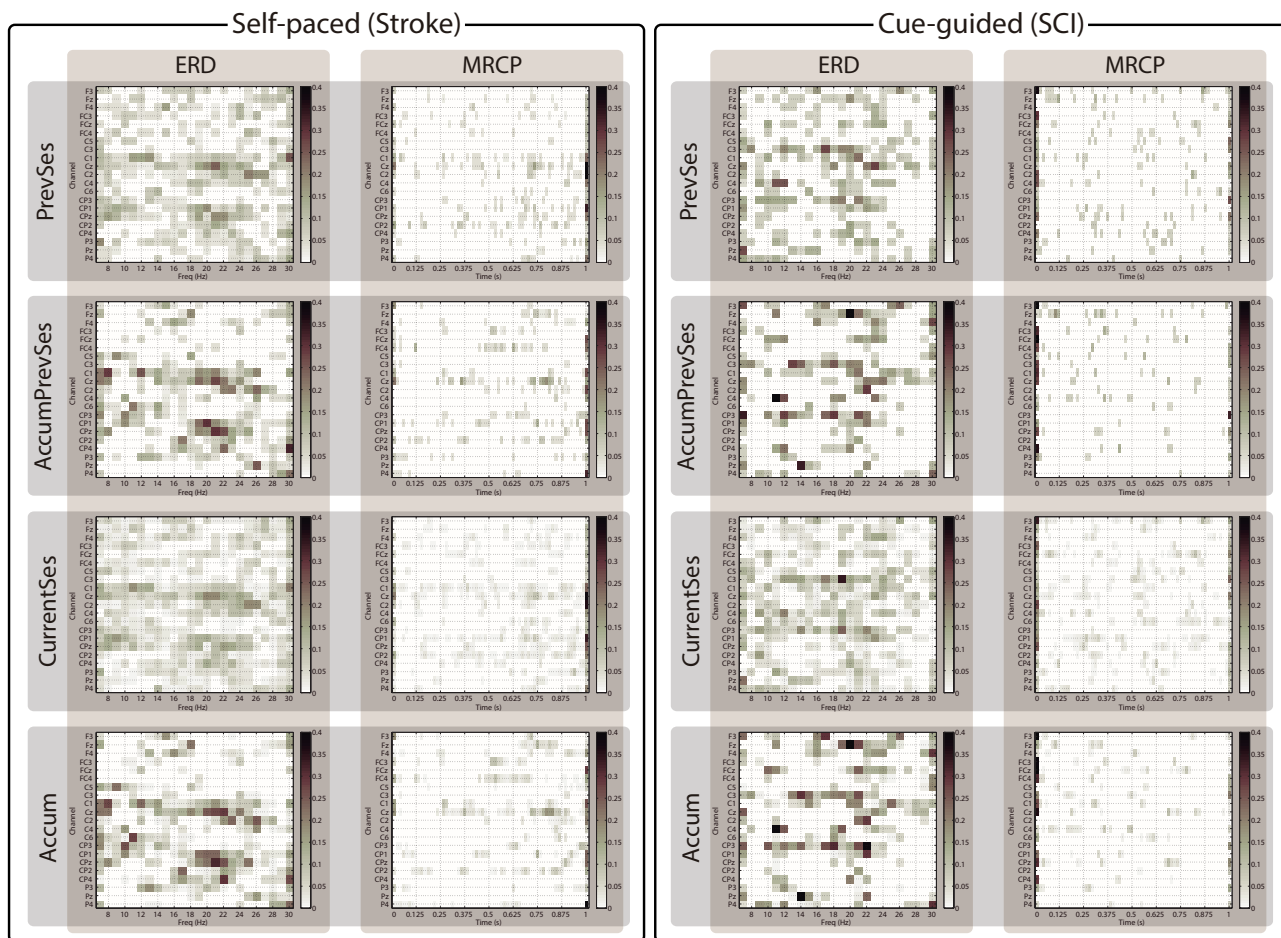
Figure 3. Features selected by SDA for the self-initiated protocol with stroke patients (left panel) and for the cue-guided protocol with SCI patients (right panel). For representation, information of all the patients (within each population) and sessions was averaged, separately for each recalibration scheme. The left/right columns within each panel depict the features extracted from ERD/MRCP activity. The four rows correspond to *PrevSes*, *AccumPrevSes*, *CurrentSes*, and *Accum* schemes. The scale of the figures represents the percentage of times that the features were chosen over all folds of all leave-one-out executions. The ERD matrices represent channel-frequency pairs, while the MRCP matrices represent channel-time pairs. As in Figure 2, the left- and right-sided electrodes were swapped in the patients that performed the task with the left hand, so that in all cases, the electrodes of the left hemisphere (odd numbers) represent the contralateral hemisphere to the arm that the patients attempted to move.

decode the movement intention with accurate temporal precision, and this is intrinsically validated by considering as TPs the classifier activations within the ±500 ms windows around the actual onsets of the movements. The average detection latencies were significantly different between the recalibration strategies ($p < 0.015$), with the post-hoc analyses showing only significant differences between *CurrentSes* and *AccumPrevSes* schemes ($p = 0.005$). No differences were found in the standard deviation of the obtained latencies with the different schemes ($p = 0.522$). Notice that the latencies of the studied schemes were

similar to the ones of the random classifier. In contrast, the standard deviations were, on average, 28% (self-paced experiment) and 24% (cue-guided experiment) smaller than the ones obtained with the random classifier (which tends to show a uniformly distributed TP profile along the windows considered to belong to the movement state).

The classification results were stable across sessions in terms of TP, FP and latencies, with no detectable trends towards better or worse performances.

Finally, Figure 5 analyzes, for *Accum* and *Cur-*

Table 4.   Results of TP, FP and latencies for each scheme

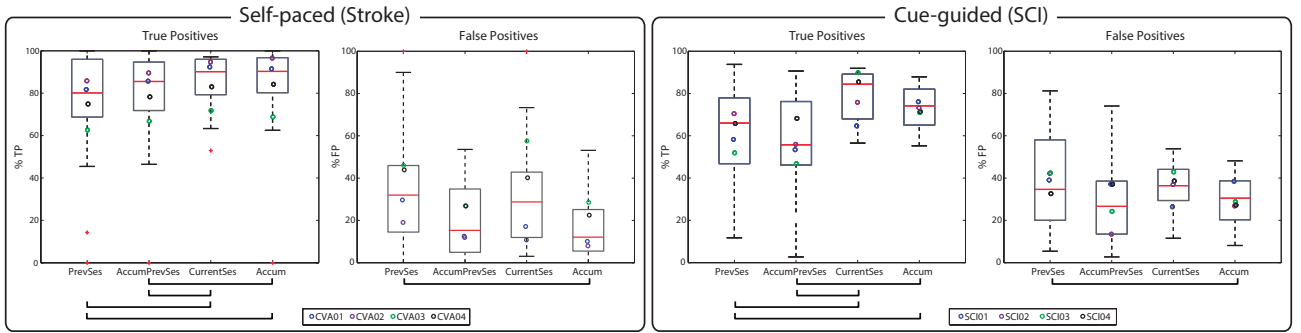| Scheme	Self-paced (Stroke)			Cue-guided (SCI)		
	TP(%)	FP(%)	Lat (s)	TP(%)	FP(%)	Lat (s)
PrevSes	76.3	34.6	-0.06±0.24	61.6	39.1	1.16±0.62
AccumPrevSes	80.1	19.5	0.02±0.22	56.1	28.0	1.28±0.58
CurrentSes	85.5	31.4	-0.07±0.21	78.9	36.2	1.10±0.60
Accum	85.3	17.3	0.01±0.21	72.9	30.3	1.22±0.57
Random classifier	34.4	49.8	-0.05±0.28	46.5	43.3	1.24±0.73
CurrentSes 40%	78.9	44.2	-0.07±0.19	91.0	69.7	1.06±0.39
Accum 40%	84.6	19.4	0.01±0.20	66.9	27.4	1.20±0.57

Figure 4.   Percentages of true positives (TP) and false positives (FP) for each calibration scheme on each group. The left panel shows the results for the protocol with self-initiated movements for stroke patients, and the right panel shows the results for the cue-based protocol with SCI patients. The boxplots on each diagram are calculated considering the performances of all the sessions of all the subjects, and on each of them is indicated the average performance of each subject.

rentSes schemes, the influence on the classifier's performance of the number of training examples considered from the session in which the BMI is tested. According to these results, the *Accum* scheme is less dependent on the number of training trials collected on the "intervention" session, and it provides the best average recalibration results when considering only 20-40% of the training examples from the classification session, *i.e.*, with around 6-12 training examples from the validation session. To facilitate the comparison with the other results based on the leave-one-out procedure, the last two rows of Table 4 represent the particular case in which *CurrentSes* and *Accum* schemes used 40% of the session data to train the decoder. In this case a higher variability could be observed, with the *CurrentSes 40%* scheme leading to substantially higher FP rates, and the *Accum 40%* being more similar to the standard *Accum* scheme.

4.   Discussion and Conclusions

This paper presents a systematic approach comparing the performance of different recalibration schemes for an EEG-based classifier of motor intentions in ecologically relevant experiments with paralyzed patients. Each scheme considered here simulates a scenario of a BMI-based rehabilitation intervention with multiple sessions, in which, for the n-th session, data from the past and/or the current ones are used to calibrate the decoder. The different recalibration schemes are evaluated offline in two datasets: an 8-sessions intervention with stroke patients using a self-paced BMI paradigm, and a 5-sessions intervention with incomplete SCI patients using a cue-guided BMI paradigm. The results highlight the importance of the data used to calibrate the movement intention BMI-decoders in order to improve their accuracy. A relevant result of this study is that significant differences can be observed between calibration approaches only using data from previous
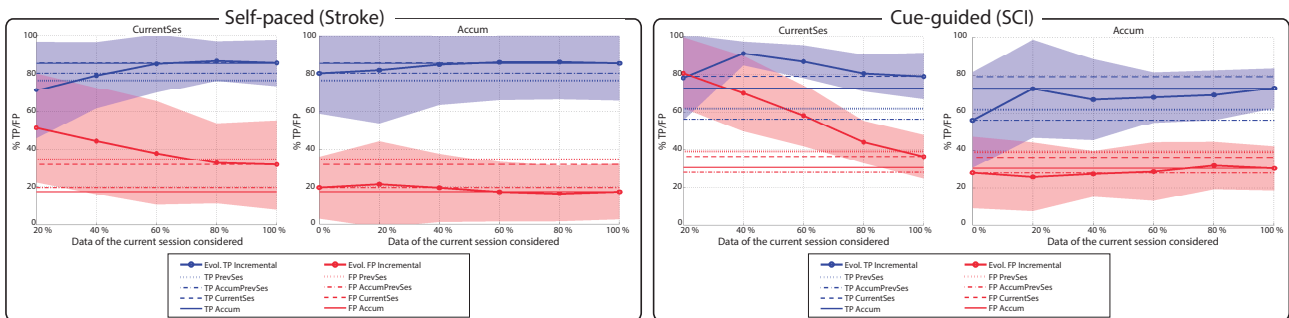
Figure 5. Changes in TPs (blue) and FPs (red) when varying the amount of data of the validation session used for training in *CurrentSes* and *Accum* schemes. The solid-thick lines and shades represent the mean±std of TPs and FPs as a function of the percentage of data from the validation session used for training: 0% (only for *Accum*), 20%, 40%, 60%, 80%, or 100%. The horizontal constant lines refer to TP/FP results using a leave-one-out procedure with each of the four schemes. Notice that, in both panels, the dashed/solid thin line corresponds to *CurrentSes*/*Accum* schemes, and hence their value is the same as for the thick line when it was evaluated with 100% of the data. In addition, the 0% in the *Accum* scheme is equivalent to the *AccumPrevSes* scheme.

calibration sessions (the most common approach in BMI interventions) and calibration approaches combining data from past sessions with data from the validation session (even if the size of the training dataset from the validation session is reduced).

With a growing interest in the integration of BMI systems for rehabilitation and motor restoration of patients with motor dysfunction, the optimization of these technologies becomes of high relevance.[33] Given the time-consuming set-up procedures required to start a BMI intervention, one of the best ways to maximize the therapy time is to reduce calibration periods. Assuming that the longer the time in which the patients receive associative feedback, the better recovery, we aim at maximizing this therapy time, while trying also to maximize the decoders' performance. Choosing optimal calibration schemes for EEG-based classifiers of motor intentions results in enhanced performances of these technologies.

According to the results here presented, EEG-based classifiers of motor intentions used along multiple sessions should base the decision on which recalibration method to use on the amount of available training data from the session in which the intervention is to be carried out. Recalibration schemes using training datasets from previous sessions combined with small sets ($\sim$ 10-12, see Fig. 5) of training examples from the intervention session result in good values of specificity and sensitivity. This is probably due to the fact that the schemes trained including data from the same session that is being tested (i.e.,

CurrentSes and *Accum*) can minimize the effects of EEG non-stationarities. Therefore, we consider that, in terms of suitability of these different schemes to real clinical interventions, carrying out large initial calibration sessions ($\sim$ 30-40 trials) in combination with smaller screening intervals preceding each BMI intervention appears to be a good solution that satisfies the compromise between time dedicated in each session to calibrate the platform and robustness of the implemented decoders of motor intentions.

One relevant finding of the present study is the similar behavior of the proposed schemes in both datasets, despite the fact that they correspond to different paradigms (self-paced vs. cue-guided) and type of patients (stroke vs. SCI). In both cases, schemes including information from the tested session (i.e., *CurrentSes* and *Accum*) provide higher TP rates than the others, while the lowest FP rates are found for accumulative schemes (i.e., *AccumPrevSes* and *Accum*). There is evidence showing that EEG signals have lower variance in power values during a motor task than during rest.[34, 35] Therefore, we speculate that the larger variance in power of the resting EEG might be better modeled when more data from the past is accumulated (*AccumPrevSes* and *Accum* schemes), while the power values of the motor condition can be better estimated using the data from the tested session (*CurrentSes* and *Accum* schemes).

Although a reduced group of patients is considered here, the consistency in the classification accuracies supports the generalizability of the results obtained. In addition, the best performing recali-

bration schemes led to classifier performances that are comparable to those presented in similar published studies.[5,6,18,21,36–38] This is remarkable because it shows, for the first time, that a single classifying strategy can match the requirements of both paradigms considered here, which are in turn the two most frequent paradigms used in BMI interventions for neurorehabilitation. On one hand, in the self-paced condition, cortical changes at the movement onset are tracked by the classifier to minimize the detection latencies, which can allow eliciting associative facilitation if used to provide a coherent afferent stimulation.[11,14] On the other hand, in the cue-guided paradigms, the cortical changes do not anticipate the patients attempts to move, and the precise identification of states of motor cortical activation is a key element to reward the patients intentions with proprioceptive feedback.[3,17] Undoubtedly, the recalibration scheme and its effects in performance is not the only relevant parameter for a successful BMI neurorehabilitation intervention. Factors impacting motor learning, such as amount and dosage of the training (e.g., total duration of the intervention, or number of sessions per week),[39] or the type of proprioceptive feedback (electrical or mechanical stimulation; above or below motor threshold; etc)[17] also play a pivotal role in the consolidation of learning, and accordingly, in the success of these therapies to maximize motor recovery. On top of that, individual factors can also affect the response of the patients to the intervention.[40,41]

Potential benefits have been demonstrated during the past years regarding the use of BMI technologies coupling motor intentions with proprioceptive feedback for motor neurorehabilitation.[1,42] Since the recent demonstration of the positive effects of BMI rehabilitation in a double-blind controlled study,[3] these results have been replicated and confirmed,[8–10] providing solid evidence of the functional improvements associated to BMI training in stroke patients. For SCI patients, BMI systems were initially devised for functional restoration of movements,[43,44] although recent trends also aim at rehabilitation of patients with incomplete[20,45] and even complete[12] SCI lesions. Notice that all the studies demonstrating functional recovery with BMIs are based on the discrimination of movement intention from resting activity (i.e., a two-class classification problem).[3,8–10,12] Works proposing systems to clas-

sify different movements from EEG have shown lower decoding performances,[46–48] and to date there is no evidence showing if decoding different movements in a BMI rehabilitation scenario can lead to higher functional improvements. However, if integrated for rehabilitation in a multiple-session paradigm, we expect that a system decoding several different movements would also benefit from recalibration.[47]

To date, insufficient efforts have been devoted to optimize BMI recalibration strategies in protocols consisting of multiple sessions, which in turn appears to be a critical element to provide robust enduring functional benefits in patients.[3,11,12] Previous studies validating multisession BMI interventions for motor neurorehabilitation have typically used fixed classifiers trained with data from an initial screening session[7,8,10,49] and in some cases intrasession-calibrated decoders have been implemented.[50] Another interesting proposed approach is to have an initial screening session to select the control features, and to use an adaptive strategy to update only the threshold used to provide feedback with data recorded during the therapeutic intervention.[3] Although this approach avoids recording screening data on each intervention session, special care has to be taken as some sorts of stimulation may induce artifacts on the recorded signals, biasing the decoder performances.[51] In a recent study performing a long term intervention (12 months) with a SCI patient, the classifier was retrained periodically and kept fixed during several sessions. Interestingly, in those periods between recalibrations, the decoding results showed a progressive decrease in performance,[44] which underlines the importance of using recent data to train the BMI decoder. Given that each proposed intervention or application presents different particularities, direct comparisons between their results are difficult, which evidences the need of meta-analyses to study the influence of calibration methodologies on BMI performances and their relationship with recovery.

In a preliminary study, we evaluated some of the schemes here presented and some subject-transfer approaches with data of SCI patients only.[27] However, we evaluated the decoding schemes on datasets that included closed-loop electrical stimulation, which might have influenced the performance of the decoder.[52] The results also showed better performances for schemes including training data of the

tested session, which agrees with a recent study that compared a movement intention decoder in 3 sessions with stroke patients with and without recalibration.[28] In addition, it was also shown that approaches combining data from other patients might not be so effective,[27] presumably due to the cortical reorganization process of these patients,[25] despite these strategies have been used in other BMI paradigms.[53–55]

An additional argument in favor of recalibrating the BMI decoders is the brain reorganization experienced by neurological patients. The data analyzed in this paper corresponds to two preliminary and short-lasting interventions, in which no significant changes in cortical activation were expected in any group of patients, which was confirmed by offline analyses (results not reported in this paper). However, two phenomenons can have an impact on the cortical patterns of the patients, which means that a decoder trained with data of a certain day only, might become non-functional some weeks or days later.[56] Firstly, the neural reorganization following a stroke or spinal cord injury causes significant changes in the brain activity of the patients in a period of few weeks.[24, 25] Secondly, the BMI interventions themselves aim at inducing a cortical reorganization to facilitate motor recovery.[3] The influence and importance of each of these two factors is difficult to quantify, and probably varies with the typology of patient and the BMI protocol carried out. However, it evidences the fact that strategies with no recalibration might result even less effective in long-lasting interventions.

The self-paced paradigm (with stroke patients) involved actual movements (or attempts of movements). Therefore, there is a possibility that motor-related artifacts had an impact on the obtained BMI performances.[57] However, this impact is estimated to be small. First, muscle artifacts are associated with increases of the EEG signal power in the beta band, which are in a different frequency band and in the opposed direction to the ERD changes learned by the classifier from the pre-movement EEG segments in the training stage (see Figure 2). Also, muscle artifacts should be better observed in the lateral regions of the ipsilateral hemisphere, while, according to the features selected, the most frequently selected channels to classify motor intentions were over central and contralateral areas (see Figure 3). Regarding post-movement low frequency components, these are

spread along the scalp, which allow spatial filtering techniques as the ones used in this study to cancel them. Moreover, these slow components are opposed to BP-related EEG changes.[58]

Finally, the comparisons between calibration schemes are dependent on the nature of the features extracted. The classifier proposed in this study makes use of both, changes in the power of cortical oscillations (to characterize the ERD) and changes in the amplitudes of the EEG slow components (BP). As a result, other EEG-based approaches to decode motor intentions based on alternative feature extraction methods (e.g., different algorithms to estimate the power,[59] or phase information of the BP[60]) may not lead to equal results when comparing different recalibration schemes.

Bibliography

1. J. J. Daly and J. R. Wolpaw, Brain-computer interfaces in neurological rehabilitation, *Lancet neurology* **7**(11) (2008) 1032–1043.
2. N. Birbaumer, A. Ramos-Murguialday and L. Cohen, Brain-computer interface in paralysis, *Curr Opin Neurol* **21**(6) (2008) 634–638.
3. A. Ramos-Murguialday, D. Broetz, M. Rea, L. Läer, O. Yilmaz, F. L. Brasil, G. Liberati, M. R. Curado, E. Garcia-Cossio, A. Vyziotis, W. Cho, M. Agostini, E. Soares, S. Soekadar, A. Caria, L. G. Cohen and N. Birbaumer, Brain-machine interface in chronic stroke rehabilitation: a controlled study, *Annals of neurology* **74**(1) (2013) 100–108.
4. I. K. Niazi, N. Jiang, O. Tiberghien, J. F. Nielsen, K. Dremstrup and D. Farina, Detection of movement intention from single-trial movement-related cortical potentials, *Journal of neural engineering* **8**(6) (2011) 066009.
5. E. López-Larraz, L. Montesano, Á. Gil-Agudo and J. Minguez, Continuous decoding of movement intention of upper limb self-initiated analytic movements from pre-movement EEG correlates, *Journal of neuroengineering and rehabilitation* **11** (2014) 153.

6. J. Ibáñez, J. I. Serrano, M. D. del Castillo, E. Monge, F. Molina, I. Alguacil and J. L. Pons, Detection of the onset of upper-limb movements based on the combined analysis of changes in the sensorimotor rhythms and slow cortical potentials, *Journal of Neural Engineering* **11**(5) (2014) 056009.

7. B. Várkuti, C. Guan, Y. Pan, K. S. Phua, K. K. Ang, C. W. K. Kuah, K. Chua, B. T. Ang, N. Birbaumer and R. Sitaram, Resting state changes in functional connectivity correlate with movement recovery for BCI and robot-assisted upper-extremity training after stroke, *Neurorehabilitation and neural repair* **27**(1) (2013) 53–62.

8. K. K. Ang, K. S. G. Chua, K. S. Phua, C. Wang, Z. Y. Chin, C. W. K. Kuah, W. Low and C. Guan, A Randomized Controlled Trial of EEG-Based Motor Imagery Brain-Computer Interface Robotic Rehabilitation for Stroke, *Clinical EEG and neuroscience* **46**(4) (2015) 310–320.

9. T. Ono, K. Shindo, K. Kawashima, N. Ota, M. Ito, T. Ota, M. Mukaino, T. Fujiwara, A. Kimura, M. Liu and J. Ushiba, Brain-computer interface with somatosensory feedback improves functional recovery from severe hemiplegia due to chronic stroke, *Frontiers in neuroengineering* **7** (2014) 19.

10. F. Pichiorri, G. Morone, M. Petti, J. Toppi, I. Pisotta, M. Molinari, S. Paolucci, M. Inghilleri, L. Astolfi, F. Cincotti and D. Mattia, Brain-computer interface boosts motor imagery practice during stroke recovery, *Annals of Neurology* **77**(5) (2015) 851–865.

11. N. Mrachacz-Kersting, N. Jiang, A. J. T. Stevenson, I. Niazi, V. Kostic, A. Pavlovic, S. Radovanovic, M. Djuric-Jovicic, F. Agosta, K. Dremstrup and D. Farina, Efficient neuroplasticity induction in chronic stroke patients by an associative brain-computer interface, *Journal of neurophysiology* **115**(3) (2016) 1410-1421.

12. A. R. C. Donati, S. Shokur, E. Morya, D. S. F. Campos, R. C. Moioli, C. M. Gitti, P. B. Augusto, S. Tripodi, C. G. Pires, G. A. Pereira, F. L. Brasil, S. Gallo, A. A. Lin, A. K. Takigami, M. A. Aratanha, S. Joshi, H. Bleuler, G. Cheng, A. Rudolph and M. A. L. Nicolelis, Long-Term Training with a Brain-Machine Interface-Based Gait Protocol Induces Partial Neurological Recovery in Paraplegic Patients, *Scientific Reports* **6** (2016) 30383.

13. P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao and K. Müller, Towards adaptive classification for BCI, *Journal of Neural Engineering* **3**(1) (2006) R13.

14. I. K. Niazi, N. Mrachacz-Kersting, N. Jiang, K. Dremstrup and D. Farina, Peripheral electrical stimulation triggered by self-paced detection of motor intention enhances motor evoked potentials, *IEEE transactions on neural systems and rehabilitation engineering* **20**(4) (2012) 595–604.

15. G. Pfurtscheller and F. H. Lopes da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles, *Clinical neurophysiology* **110**(11) (1999) 1842–1857.

16. H. Shibasaki and M. Hallett, What is the Bereitschaftspotential?, *Clinical neurophysiology* **117**(11) (2006) 2341–2356.

17. A. Ramos-Murguialday, M. Schürholz, V. Caggiano, M. Wildgruber, A. Caria, E. M. Hammer, S. Halder and N. Birbaumer, Proprioceptive feedback and brain computer interface (BCI) based neuroprostheses, *PLoS ONE* **7**(10) (2012) e47048.

18. R. Xu, N. Jiang, N. Mrachacz-Kersting, C. Lin, G. Asin, J. Moreno, J. L. Pons, K. Dremstrup and D. Farina, A Closed-Loop Brain-Computer Interface Triggering an Active Ankle-Foot Orthosis for Inducing Cortical Neural Plasticity, *IEEE Transactions on Biomedical Engineering* **61**(7) (2014) 2092-2101.

19. J. Ibáñez, J. I. Serrano, M. D. del Castillo, E. Monge, F. Molina, F. M. Rivas and I. Alguacil, Detection of the Onset of Voluntary Movements Based on the Combination of ERD and BP Cortical Patterns, *Proceedings of the 2nd International Conference on NeuroRehabilitation (ICNR2014)*, (2014), 437–446.

20. E. López-Larraz, F. Trincado-Alonso, V. Rajasekaran, A. J. Del-Ama, J. Aranda, J. Minguez, Á. Gil-Agudo and L. Montesano, Control of an ambulatory exoskeleton with a brain-machine interface for spinal cord injury gait rehabilitation, *Frontiers in Neuroscience* **10** (2016) 359.

21. J. Ibáñez, E. Monge-Pereira, F. Molina-Rueda, J. I. Serrano, M. D. del Castillo, A. Cuesta-Gómez, M. Carratalá-Tejada, R. Cano-de-la Cuerda, I. M. Alguacil-Diego, J. C. Miangolarra-Page and J. L. Pons, Low Latency Estimation of Motor Intentions to Assist Reaching Movements along Multiple Sessions in Chronic Stroke Patients: A Feasibility Study, *Frontiers in Neuroscience* **11** (2017) 126.

22. I. K. Niazi, N. Jiang, M. Jochumsen, J. F. Nielsen, K. Dremstrup and D. Farina, Detection of movement-related cortical potentials based on subject-independent training, *Medical and Biological Engineering and Computing* **51**(5) (2013) 507–512.

23. J. J. Daly, Y. Fang, E. M. Perepezko, V. Siemionow and G. H. Yue, Prolonged cognitive planning time, elevated cognitive effort, and relationship to coordination and motor control following stroke, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **14**(2) (2006) 168–171.

24. M. Stępień, J. Conradi, G. Waterstraat, F. U. Hohlefeld, G. Curio and V. V. Nikulin, Event-related desynchronization of sensorimotor EEG rhythms in hemiparetic patients with acute stroke, *Neuroscience letters* **488**(1) (2011) 17–21.

25. E. López-Larraz, L. Montesano, Á. Gil-Agudo, J. Minguez and A. Oliviero, Evolution of EEG motor rhythms after spinal cord injury: a longitudinal study, *PLoS ONE* **10**(7) (2015) e0131759.

26. M. Arvaneh, C. Guan, K. K. Ang and C. Quek, Omitting the intra-session calibration in EEG-based

brain computer interface used for stroke rehabilitation, *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (2012) 4124–4127.

27. E. López-Larraz, F. Trincado-Alonso and L. Montesano, Brain-machine interfaces for motor rehabilitation: is recalibration important?, *14th International Conference on Rehabilitation Robotics (ICORR)*, (2015), 223–228.

28. A. I. Sburlea, L. Montesano and J. Minguez, Intersession adaptation of the EEG-based detector of self-paced walking intention in stroke patients, *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (2015) 498–501.

29. L. Clemmensen, T. Hastie, D. Witten and B. Ersbøll, Sparse Discriminant Analysis, *Technometrics* **53**(4) (2011) 406–413.

30. Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang and A. Cichocki, Aggregation of sparse linear discriminant analyses for event-related potential classification in brain-computer interface, *International journal of neural systems* **24**(1) (2014) 1450003.

31. A. I. Sburlea, L. Montesano, R. C. de la Cuerda, I. M. Alguacil Diego, J. C. Miangolarra-Page and J. Minguez, Detecting intention to walk in stroke patients from pre-movement EEG correlates, *Journal of NeuroEngineering and Rehabilitation* **12**(1) (2015) 113.

32. W.-Y. Hsu, Continuous EEG Signal Analysis for Asynchronous BCI Application, *International Journal of Neural Systems* **21**(04) (2011) 335–350.

33. R. Rupp, Challenges in clinical applications of brain computer interfaces in individuals with spinal cord injury, *Frontiers in neuroengineering* **7** (2014) 38.

34. B. Graimann, J. E. Huggins, S. P. Levine and G. Pfurtscheller, Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data, *Clinical neurophysiology* **113**(1) (2002) 43–47.

35. B. Graimann and G. Pfurtscheller, Quantification and visualization of event-related changes in oscillatory brain activity in the time-frequency domain, *Progress in brain research* **159** (2006) 79–97.

36. R. Xu, N. Jiang, C. Lin, N. Mrachacz-kersting, K. Dremstrup and D. Farina, Enhanced low-latency detection of motor intention from EEG for closed-loop brain-computer interface applications, *IEEE Transactions on biomedical engineering* **61**(2) (2014) 288–296.

37. E. Abou Zeid and T. Chau, Electrode Fusion for the Prediction of Self-Initiated Fine Movements from Single-Trial Readiness Potentials, *International Journal of Neural Systems* **25**(04) (2015) 1550014.

38. E. López-Larraz, A. M. Ray, T. C. Figueiredo, C. Bibián, N. Birbaumer and A. Ramos-Murguialday, Stroke lesion location influences the decoding of movement intention from EEG, *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (2017) 3065–3068.

39. R. Schmidt and T. Lee, *Motor Control and Learning: A Behavioral Emphasis*, (2011).

40. S. H. George, M. H. Rafiei, L. Gauthier, A. Borstad, J. A. Buford and H. Adeli, Computer-aided prediction of extent of motor recovery following constraint-induced movement therapy in chronic stroke, *Behavioural Brain Research* **329** (2017) 191–199.

41. S. H. George, M. H. Rafiei, A. Borstad, H. Adeli and L. V. Gauthier, Gross motor ability predicts response to upper extremity rehabilitation in chronic stroke, *Behavioural Brain Research* **333** (2017) 314–322.

42. L. E. H. V. Dokkum, T. Ward and I. Laffont, Brain computer interfaces for neurorehabilitation–its current status as a rehabilitation strategy post-stroke, *Annals of Physical and Rehabilitation Medicine* **58**(1) (2015) 3–8.

43. C. Enzinger, S. Ropele, F. Fazekas, M. Loitfelder, F. Gorani, T. Seifert, G. Reiter, C. Neuper, G. Pfurtscheller and G. Müller-Putz, Brain motor system function in a patient with complete spinal cord injury following extensive brain-computer interface training, *Experimental brain research* **190**(2) (2008) 215–223.

44. M. Rohm, M. Schneiders, C. Müller, A. Kreilinger, V. Kaiser, G. R. Müller-Putz and R. Rupp, Hybrid brain-computer interfaces and hybrid neuroprostheses for restoration of upper limb functions in individuals with high-level spinal cord injury, *Artificial intelligence in medicine* **59**(2) (2013) 133–142.

45. C. E. King, P. T. Wang, C. M. McCrimmon, C. C. Chou, A. H. Do and Z. Nenadic, The feasibility of a brain-computer interface functional electrical stimulation system for the restoration of overground walking after paraplegia, *Journal of NeuroEngineering and Rehabilitation* **12**(1) (2015) 80.

46. J. Ibáñez, J. I. Serrano, M. D. del Castillo, J. Minguez and J. L. Pons, Predictive classification of self-paced upper-limb analytical movements with EEG, *Medical & biological engineering & computing* **53**(11) (2015) 1201-1210.

47. F. Shiman, E. López-Larraz, A. Sarasola-Sanz, N. Irastorza-Landa, M. Spueler, N. Birbaumer and A. Ramos-Murguialday, Classification of different reaching movements from the same limb using EEG, *Journal of Neural Engineering* **14** (2017) 046018.

48. P. Ofner, A. Schwarz, J. Pereira and G. R. Müller-Putz, Upper limb movements can be decoded from the time-domain of low-frequency EEG, *PLoS ONE* **12**(8) (2017).

49. J. J. Daly, R. Cheng, J. Rogers, K. Litinas, K. Hrovat and M. Dohring, Feasibility of a new application of noninvasive Brain Computer Interface (BCI): a case study of training for recovery of volitional motor control after stroke, *Journal of neurologic physical therapy* **33**(4) (2009) 203–211.

50. B. M. Young, Z. Nigogosyan, L. M. Walton, A. Remsik, J. Song, V. A. Nair, M. E. Tyler, D. F. Edwards, K. Caldera, J. A. Sattin, J. C. Williams and V. Prabhakaran, Dose-response relationships using braincomputer interface technology impact stroke rehabilitation, *Frontiers in Human Neuroscience* **9** (2015) 1–14.

51. A. Walter, A. Ramos Murguialday, M. Spüler, G. Naros, M. T. Leão, A. Gharabaghi, W. Rosenstiel, N. Birbaumer and M. Bogdan, Coupling BCI and cortical stimulation for brain-state-dependent stimulation: methods for spectral estimation in the presence of stimulation after-effects, *Frontiers in neural circuits* **6** (2012) 87.

52. D. Zhang, F. Xu, H. Xu, P. B. Shull and X. Zhu, Quantifying Different Tactile Sensations Evoked by Cutaneous Electrical Stimulation Using Electroencephalography Features, *International Journal of Neural Systems* **26**(2) (2016) 1650006.

53. M. Xu, J. Liu, L. Chen, H. Qi, F. He, P. Zhou, B. Wan and D. Ming, Incorporation of Inter-Subject Information to Improve the Accuracy of Subject-Specific P300 Classifiers, *International Journal of Neural Systems* **26**(3) (2016) 1650010.

54. E. López-Larraz, I. Iturrate, C. Escolano, I. García, L. Montesano and J. Minguez, Single-trial classification of feedback potentials within neurofeedback training with an EEG brain-computer interface, *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (2011), 4596–4599.

55. E. López-Larraz, M. Creatura, I. Iturrate, L. Montesano and J. Minguez, EEG single-trial classification of visual, auditive and vibratory feedback potentials in brain-computer interfaces, *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (2011), 4231–4234.

56. Y. Liu, Q. Zhao and L. Zhang, Uncorrelated multiway discriminant analysis for motor imagery EEG classification, *International journal of neural systems* **25**(4) (2015) 1550013.

57. E. López-Larraz, C. Bibián, N. Birbaumer and A. Ramos-Murguialday, Influence of artifacts on movement intention decoding from EEG activity in severely paralyzed stroke patients, *15th International Conference on Rehabilitation Robotics (ICORR)*, (2017) 901–906.

58. H. Shibasaki, G. Barrett, E. Halliday and a. M. Halliday, Components of the movement-related cortical potential and their scalp topography, *Electroencephalography and clinical neurophysiology* **49**(3) (1980) 213–226.

59. W.Y. Hsu, Assembling A Multi-Feature EEG Classifier for Left-Right Motor Imagery Data Using Wavelet-Based Fuzzy Approximate Entropy for Improved Accuracy, *International journal of neural systems* **25**(8) (2015) 1550037.

60. A. I. Sburlea, L. Montesano and J. Minguez, Advantages of EEG phase patterns for the detection of gait intention in healthy and stroke subjects, *Journal of Neural Engineering* **14**(3) (2017) 036004.