Brain-machine interfaces for motor rehabilitation: Is recalibration important?

Eduardo López-Larraz*†, Fernando Trincado-Alonso‡, Luis Montesano*†
*Departamento de Informática e Ingeniería de Sistemas, University of Zaragoza, Spain
†Instituto de Investigación en Ingeniería de Aragón, Zaragoza, Spain
‡Biomechanics and Technical Aids Unit, National Hospital for Spinal Cord Injury, Toledo, Spain.

***Abstract*—Brain-machine interfaces (BMI) allow to decode motor commands from paralyzed patients' brains and use those commands with a rehabilitative or assistive purpose. However, brain non-stationarities can affect BMI performance over time in multi-session interventions. The amount and type of data used for calibration may play an important role on the posterior decoding performance. This paper studies six different schemes for BMI calibration, considering subject-specific and subject-transfer scenarios. Data from a five-session rehabilitation intervention with four spinal cord injury patients is used to evaluate the decoding performance of the six proposed schemes. Our results show that recording some data at the beginning of each new session to recalibrate the BMI has a positive effect, although this effect is not achieved if we do not record enough number of trials. In addition, for subject-transfer approaches it is possible to achieve similar performances to those of subject-specific approaches for some subjects, but for others, generalization is not possible. These findings constitute a step forward towards the implantation of BMI for multiple-session rehabilitation therapies.**

I. Introduction

Rehabilitation of neuro-motor disabilities is, in the recent years, taking advantage of robotic and orthotic devices to improve the therapeutic outcomes. Different studies have successfully applied such devices for rehabilitation of paralyzed patients with spinal cord injury (SCI) or stroke [1], [2], [3]. More specifically, rehabilitation systems triggered by neural commands are interesting as they have the potential to induce neural plasticity [4]. Brain-machine interfaces (BMI) allow the real-time decoding of electroencephalographic (EEG) signals, permitting an additional channel of interaction for paralyzed patients [5], [6], [7]. Indeed, recent studies have shown the feasibility of BMI to trigger rehabilitative devices, achieving very promising results [8], [9], [10].

One of the challenges for the outright implantation of BMI technology in rehabilitation is the optimization of the decoders that translate human thoughts into computer commands. Indeed, calibration is one of the most important factors for the applicability of BMI in daily rehabilitation practice [11]. Pilot BMI studies are generally performed with healthy subjects on lab environments, and lots of data are recorded to evaluate decoding performances. However, when involving highly dependent patients it is not convenient to execute long experimentation sessions. Moreover, most of the session time should be used for rehabilitative purpose, and not for equipment donning or system calibration. Although several studies have conducted multi-session BMI interventions for motor rehabilitation, there is still no consensus about how to calibrate the BMI to improve the decoding performance. Some different approaches used are: an initial calibration for all the therapy [12]; calibrations at the beginning of each session [10], [13]; and recalibrations combining data from previous sessions [8], [14].

In this line, effort is being conducted towards the design of systems that require a small calibration time and work with high performance, providing also stability in multi-session studies. The evaluation of BMI generalization across sessions has been studied for classification of event related potentials [15], [16], [17]. For BMI paradigms to decode motor commands, several works have studied the reduction of calibration time in new sessions by considering information from other subjects [18] or from past sessions [14], [19]. However, these two generalization approaches (across sessions or subjects) present additional difficulties for patients suffering from neurological injuries such as stroke or SCI. These patients show high variabilities in motor brain activations, even for those with similar lesions; additionally, they suffer from a temporal evolution of brain activations due to the neuroplastic changes following their lesions [20], [21]. Hence, we hypothesize that generalization across sessions might be more effective than across subjects for multi-session BMI interventions in this population.

This paper evaluates the influence of the data used to train a BMI decoder on its decoding accuracy, in a multi-session rehabilitation intervention. For this analysis, we used data from a pilot study of BMI-based hand rehabilitation with SCI patients. The protocol consisted of five sessions, in which the patients were asked to attempt to move their paralyzed hand, and were stimulated when the BMI decoded the intention of motion. On the first session, a long screening was performed to calibrate the BMI, while on the remaining ones, short screenings were conducted for BMI recalibration. On each session, after collecting the training data, the patients performed some test blocks. In order to evaluate the influence that the training dataset has on the decoding performance, this paper compares four decoding schemes relying on subject-specific data, and two schemes that combine data from other subjects. For each scheme, we evaluated, in a pseudo-online manner, how the BMI performance would have been on the test blocks.

978-1-4799-1808-9/15/$31.00 ©2015 IEEE

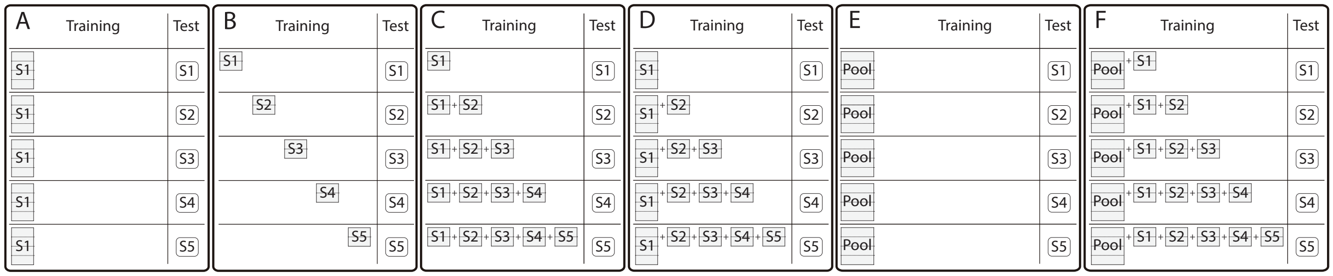

Fig. 1. Decoding schemes proposed. On each panel (A-F), the datasets used for training and test on each session are specified. Panels A-D correspond to subject-specific schemes, and panels E-F to subject transfer schemes. The left part of each panel indicates the blocks of screenings used for training. The right part of the panels indicates the test sets of each session (i.e., the test blocks of the corresponding session). A: Initial session calibration. B: Current session short-calibration. C: Short recalibration. D: Complete initial plus short-recalibration. E: Fixed pool of subjects. F: Initial pool plus short recalibration.

II. Methods

The data analyzed on this work was acquired during a pilot study for hand rehabilitation of spinal cord injury patients. The patients performed an intervention that used a BMI to trigger a combined feedback, based on functional electrical stimulation (FES) of the hand, and virtual reality showing a hand closing on a screen.

A. Patients

Four patients with incomplete tetraplegia participated in the study. All of them were in a subacute state, 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 I. A clinician evaluated the patients' condition before the study in order to select the most appropriate hand for the intervention.

TABLE I. Details of 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	10	C5, ASIA C	Male	Right	Left
P3	36	7	C5, ASIA B	Male	Right	Right
P4	55	4	C4, ASIA C	Male	Right	Right

B. Data acquisition

EEG was recorded from 32 electrodes placed at AFz, FC3, FCz, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, FP1, FP2, F7, F3, Fz, F4, F8, T7, T8, P7, P3, Pz, P4, P8, O1 and O2 (according to the international 10/10 system). The ground and reference electrodes were placed on FPz and on the left earlobe, respectively. 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.

C. Experimental design

The experimental protocol consisted of five sessions, performed within a maximum time of 10 days. During the sessions, the patients were seated on their wheelchair, facing a computer screen, and with the FES electrodes attached on one of their arms. The sessions consisted of screening blocks and closed-loop test blocks. On the first session, the patients were asked to perform 4 screening and 2 test blocks, while on the remaining four sessions, they were asked to perform 2 screening and 4 test blocks. The patients were informed about the type of block before its start.

Each screening block was composed of 20 trials, which included a rest interval (with random duration between 4 and 7 seconds), and a movement attempt (MA) interval (3 seconds). The test blocks also consisted of 20 trials, with a rest interval (10 seconds), and a MA interval (3 seconds). Visual cues were used to denote each interval. For both types of blocks, the patients were instructed to stay relaxed during the rest interval, and to attempt to grasp one of their hands (see Table I) during the MA interval. During the screening blocks, the patients did not receive any kind of feedback; during the test blocks, if one trial was correctly decoded by the BMI, the patient was stimulated by means of a FES activation and virtual reality. On each intervention session, the BMI decoder was calibrated after recording the screening blocks, using the trials of all the screening blocks of that subject recorded to that moment.

D. Feature extraction

The BMI decoder used different configurations of training datasets to distinguish between the *rest* and *movement attempt* brain states. One-second time windows were extracted from the training trials in the interval [-4, -1] s to model the *rest* class, and from the interval [0, 3] s for the *movement attempt* class[1]. The windows were slided with an overlap of 0.75 s.

Two types of features were extracted: frequency, for the event related desynchronization (ERD) [22]; and temporal, for the motor related cortical potentials (MRCP) [23]. The ERD features were extracted from 15 channels, including fronto-central (FCx), central (Cx), and centro-parietal (CPx) locations. A Laplacian filter was applied, and an autoregressive model was used to obtain the power values in the frequency range [7-30] Hz. The MRCP features were extracted from 13 channels, (i.e., the same channel set as for ERD, but removing C5 and C6 locations). A

[1]Notice that $t = 0$ corresponds to the time instant where the patients were indicated to start the movement attempt.

common average reference was applied, and the signals were downsampled to 64 Hz and filtered in the range [0.1-1] Hz. In total, 1192 features were extracted for each one-second window.

Sparse discriminant analysis [24] was used to select the most discriminative features, and as online classifier. This technique, which is a modified version of linear discriminant analysis (LDA) with a sparseness criterion, has successfully been used to decode motor brain commands both in healthy subjects and in SCI patients [7].

E. Calibration schemes

We evaluated subject-specific decoding with four different calibration schemes, and subject-transfer with two additional schemes.

1) Subject-specific decoding: Four subject-specific decoding schemes were proposed. All of them were tested for each subject separately. For each scheme, the BMI decoder was calibrated using different configurations of datasets, and tested on the test blocks of each session (Figure 1):

A) **Initial session calibration** (Fig 1A): Initial long calibration that is used for all the sessions. The training dataset consists of the four screening blocks of the first session, and it is used to classify the five sessions.

B) **Current session short-calibration** (Fig 1B): Short calibrations with data recorded every session. On each session, the training dataset consists of the two screening blocks recorded on that session (as the first session included four blocks, the two first ones are selected).

C) **Short recalibration** (Fig 1C): Incremental strategy that combines short recalibrations on each session. On the first session, the training dataset is initialized as the two first screening blocks; on the subsequent sessions, their two screening blocks are added to the training dataset of the previous session.

D) **Complete initial plus short-recalibration** (Fig 1D): Incremental strategy that combines an initial long calibration and short recalibrations on the remaining sessions. On the first session, the training dataset is initialized with the four screening blocks; on the subsequent sessions, their two screening blocks are added to the training dataset of the previous session. Notice that this was the scheme used during the intervention.

In addition, for the schemes that included short calibrations/recalibrations (i.e., B, C, and D), we tested their performance when considering the two screening blocks of each session or just one of them (the one recorded first, in each case). This allowed to compare the differences in these schemes when using 20 or 40 trials from each session. Hence, they will be referred to as B1, B2, C1, C2, D1, and D2, according to the number of screening blocks from each session considered for their calibration.

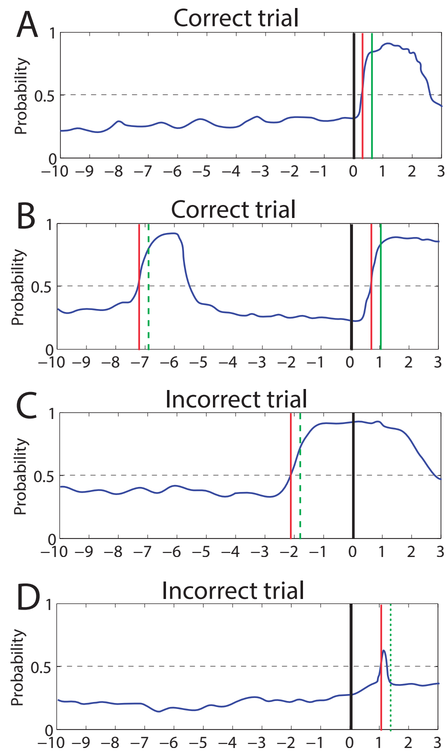

Fig. 2. Examples of correct and incorrect trials. The x-axes represent the time in seconds, while the y-axes represent the decoder probability. The horizontal, dashed, black line corresponds to the classifier threshold, and the vertical, solid, black line denotes the time of presentation of the MA cue. The time instants where the classifier changes from *rest* to *movement attempt* class are represented with the red vertical lines. If the classifier continues in *movement attempt* class during five consecutive sliding windows, a trigger is generated. This trigger is considered as valid if it is within MA interval (green solid lines, trials A-B), and not considered if it is in rest interval (green dashed lines, trials B-C). If the classifier changes from *movement attempt* to *rest* class before the fifth window, the trigger is not generated (dotted green lines, trial D).

2) Subject-transfer decoding: Two subject-transfer decoding schemes were proposed. The schemes were tested on each of the four subjects separately, by combining the data from the remaining three subjects to train the decoder.

E) **Fixed pool of subjects**: Simulation of a database of patients that is used for all the sessions. The training dataset consists of all the screening blocks from three subjects, and it is used to classify the five sessions of the remaining subject.

F) **Initial pool plus short-recalibration**: Initial database of patients that is updated every session with data of the studied patient. On the first session, the training dataset is initialized with the pool of subjects and the two first screening blocks of the studied patient; on the subsequent sessions, their two screening blocks are added to the training dataset of the previous session.

F. Decoding evaluation

All the schemes were tested simulating the online operation of the system. On each case, the BMI decoder

TABLE II. PERFORMANCES OF SUBJECT-SPECIFIC DECODING SCHEMES

Decoding Scheme	S1	S2	S3	S4	S5	Avg
Initial session calibration (A)	$76.9 \pm 17.7\%$	$63.8 \pm 35.4\%$	$52.6 \pm 40.0\%$	$42.8 \pm 40.5\%$	$56.6 \pm 31.8\%$	$\mathbf{58.5 \pm 12.7\%}$
Current session short-calibr. (B1)	$56.3 \pm 22.4\%$	$66.3 \pm 22.4\%$	$76.3 \pm 14.4\%$	$67.0 \pm 7.4\%$	$67.7 \pm 25.6\%$	$\mathbf{66.7 \pm 7.1\%}$
Current session short-calibr. (B2)	$75.6 \pm 9.7\%$	$72.5 \pm 23.2\%$	$70.7 \pm 14.2\%$	$88.1 \pm 12.1\%$	$67.2 \pm 19.9\%$	$\mathbf{74.8 \pm 8.0\%}$
Short recalibration (C1)	$56.3 \pm 22.4\%$	$59.1 \pm 22.1\%$	$63.6 \pm 16.3\%$	$58.3 \pm 33.5\%$	$50.7 \pm 29.6\%$	$\mathbf{57.6 \pm 4.7\%}$
Short recalibration (C2)	$75.6 \pm 9.7\%$	$78.5 \pm 19.3\%$	$77.7 \pm 8.1\%$	$75.9 \pm 20.0\%$	$57.2 \pm 29.8\%$	$\mathbf{73.0 \pm 8.9\%}$
Complete init. + short-recalibr. (D1)	$76.9 \pm 17.7\%$	$64.4 \pm 19.4\%$	$71.5 \pm 14.6\%$	$56.9 \pm 32.0\%$	$50.4 \pm 30.5\%$	$\mathbf{64.0 \pm 10.7\%}$
Complete init. + short-recalibr. (D2)	$76.9 \pm 17.7\%$	$88.2 \pm 8.0\%$	$82.9 \pm 13.2\%$	$77.3 \pm 17.9\%$	$62.9 \pm 36.5\%$	$\mathbf{77.6 \pm 9.4\%}$

was trained with the corresponding datasets, and the test blocks were evaluated with a one-second sliding window. The decoding performance of each scheme was evaluated by computing the percentage of test trials that were correctly decoded. As we used a cue-based protocol to indicate the patients when starting the attempt of movement, we defined the correct detections as the BMI triggers generated after the presentation of the cue ($t = 0$). Figure 2 shows four examples of correct and incorrect trials. To produce a correct trigger, the BMI had to classify five consecutive windows as *movement attempt* class, after being in *rest* class (Fig. 2A-B). This was to ensure that the triggers were generated by consistent brain activations and not by spurious detections, as in [9]. If one trigger was generated before the cue, it was considered as invalid (Fig. 2B-C). If the BMI output was not maintained in *movement attempt* class during five windows, the trigger was not generated (Fig. 2D).

III. RESULTS

A. Subject-specific decoding

Table II shows the decoding performances (mean±std) of each scheme on each session, averaged for the four subjects. Notice that schemes A, D1, and D2; schemes B1 and C1; and schemes B2 and C2 present the same accuracies for the first session as, for that session, they were trained with the same dataset (cf. Figure 1). Scheme A (first session calibration) achieved high performance only for the first session (76.9%), but suffered from a high drop on the subsequent sessions (53.95% for the average of sessions 2 to 5). This explains the higher standard deviation of scheme A (12.7) compared with the other schemes. On the other hand, all the remaining schemes included some training data from the classified session to calibrate the decoder, which resulted in lower standard deviations, but not always in higher accuracies. Scheme C1 (short recalibration with 1 screening block) presented the lowest decoding accuracy (57.6%), while scheme D2 (complete init. plus short recalibration with 2 screening blocks) presented the highest (77.6%).

Figure 3 displays the boxplots of the decoding accuracies of each scheme, including all the subjects and sessions. We compared the data distributions of the schemes that included one and two screening blocks to measure the influence of adding to the training dataset more trials recorded on the current session. We found that the three schemes provided significantly higher decoding accuracies when were trained with two screenings instead of with one (3 Wilcoxon paired tests, comparing B2 vs B1; C2 vs C1; D2 vs D1; $p < 0.05$, FDR corrected for multiple comparisons). In addition, we evaluated the influence of the scheme used for training by pairwise comparisons between the accuracy distributions of A, B2, C2, and D2 schemes (6 Wilcoxon paired tests, $p < 0.05$, FDR corrected for multiple comparisons). The scheme D2 provided significantly higher accuracies than scheme A, although no other comparison provided significant results.

B. Subject-transfer decoding

Table III shows the decoding performances (mean±std) of the four subject-specific (i.e., A, B2, C2, D2 schemes) and the two subject-transfer schemes on each patient, averaged for the five sessions. Note that, in this case, we averaged all the sessions of each patient to evaluate how each scheme worked for each of the participants. Hence, standard deviations of "Avg" row correspond to the four patients, unlike in Table II, where standard deviations of column "Avg" corresponded to the five sessions. While scheme A had a high variability between patients (std of 24.3), schemes B2, C2, and D2 were considerably more stable (std of 5.2, 9.6, and 15.4, respectively). The subject-transfer schemes provided, on average, worse decoding performances than subject-specific ones (53.8% and 53.0% for schemes E and F, respectively). However, this is due to the fact that these schemes did not decode the attempts of movement from patient P4 (0% of decoded trials for scheme E, and 3.4% for scheme F). The average decoding accuracy for the remaining three patients (i.e., P1, P2, and P3) was $71.7 \pm 7.3\%$ and $69.5 \pm 3.3\%$ for schemes E and F, respectively. Furthermore, the performances of these schemes were rather stables across sessions for the three patients, and indeed, the addition of data from previous days in scheme F did not have a positive effect over sessions.

IV. DISCUSSION AND CONCLUSIONS

In this paper we studied the impact of different calibration strategies on the performance of a BMI decoder for motion intention in a rehabilitation context. We compared four subject-specific and two subject-transfer calibration schemes, and evaluated their performances using the data recorded during a five-session intervention for hand rehabilitation with four spinal cord injury patients. Our results revealed that the training dataset used to calibrate the BMI has a big impact on the decoding accuracy. Indeed, we observed that recording some data at the beginning of each session to recalibrate the BMI has a positive effect, although this effect is not achieved if we do not record enough number of trials. Furthermore, our results suggest that when using subject-transfer approaches it might be possible to achieve

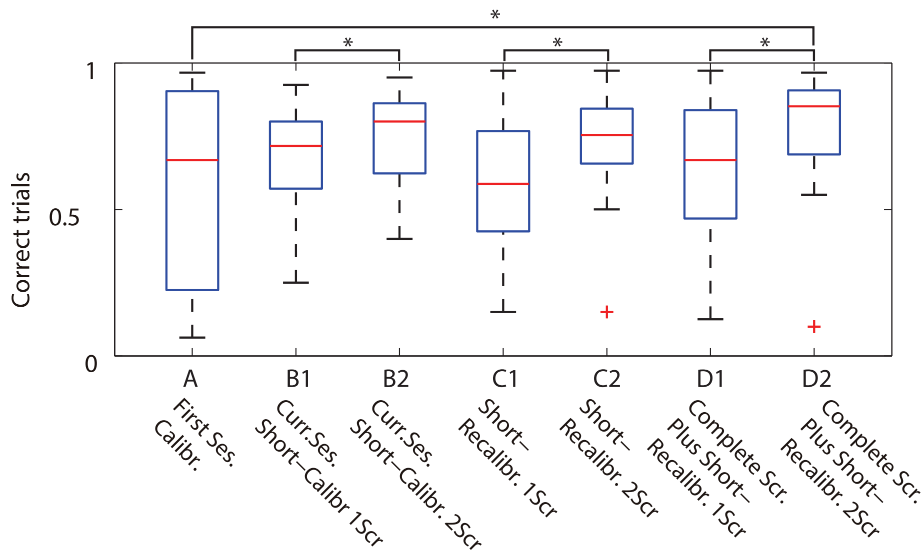

Fig. 3. Distribution of decoding accuracies for the subject-specific schemes. Each boxplot is built considering the accuracies of all the subjects and sessions. On each box, the red line represents the distribution median, while the box edges are the 25th and 75th percentiles. The whiskers are extended to include the most extreme data points not considered outliers–which are represented as the red '+'. The asterisks mark the individual comparisons which reached statistical significance ($p < 0.05$).

TABLE III. Comparison between subject-specific and subject-transfer decoding schemes

Patient ID	First Ses. Calibr. (A)	Curr. Ses. Short-Calibr. (B2)	Short-Recalibr. (C2)	Complete Scr. Plus Short-Recalibr. (D2)	Fixed Pool of Subs. (E)	Init. Pool Plus Short-Recalibr. (F)
P1	$27.5 \pm 20.1\%$	$75.5 \pm 19.2\%$	$62.0 \pm 27.2\%$	$54.5 \pm 28.2\%$	$80.0 \pm 11.0\%$	$72.5 \pm 16.6\%$
P2	$71.4 \pm 30.5\%$	$80.1 \pm 15.0\%$	$85.5 \pm 11.1\%$	$86.5 \pm 10.8\%$	$66.3 \pm 19.7\%$	$70.0 \pm 19.2\%$
P3	$52.3 \pm 35.3\%$	$76.0 \pm 15.2\%$	$71.3 \pm 12.4\%$	$85.5 \pm 4.9\%$	$68.8 \pm 27.5\%$	$66.0 \pm 23.0\%$
P4	$82.9 \pm 15.1\%$	$67.7 \pm 19.7\%$	$73.2 \pm 17.7\%$	$83.9 \pm 13.3\%$	$0.0 \pm 0.0\%$	$3.5 \pm 4.1\%$
Avg	$58.5 \pm 24.3\%$	$74.8 \pm 5.2\%$	$73.0 \pm 9.6\%$	$77.6 \pm 15.4\%$	$53.8 \pm 36.3\%$	$53.0 \pm 33.1\%$

similar performances to those of subject-specific approaches for certain subjects, although not for all of them.

The number of training trials recorded on the session to be tested has a significant impact on the accuracy. For the three schemes in which we compared the calibration/recalibration using one or two screening blocks (i.e., schemes B, C, and D), we observed significantly higher accuracies when using two blocks. Based on this analysis, we can conclude that spending more time at the beginning of each session to calibrate the BMI has a positive effect on the BMI accuracy. However, it is important to find a trade-off between the time spent for calibration and the desired BMI performance. Taking into account the results we obtained in these analyses, we consider that recording about 40 trials on each session (around 6 minutes of recording) can be sufficient. Indeed, the results obtained for scheme B2 (calibration with 40 trials recorded on the tested session) provided very similar accuracy (74.8%) to scheme A (calibration with 80 trials recorded on the first session) evaluated on the test blocks of the first session (76.9%). Hence, in the authors' opinion, scheme B2 could be the most suitable for future interventions, as it provided a good balance between time and accuracy.

As we hypothesized, generalization across sessions provided better results than across different subjects. For three of the patients, we achieved around 70% of correct trials with subject-transfer schemes, even when not considering for calibration any information of the studied patient (scheme E). However, for patient P4 we were not able to decode his motor intention, either when considering just data from the other patients or when recalibrating the pool of subjects with P4's data. This phenomenon is very likely due to differences in the brain activations of the patients, which is normal in pathologies such as SCI [20]. Although this may be solved by building a pool of subjects with a larger number of patients, further research is needed to verify this hypothesis.

It is important to note that the results presented in this work correspond to a preliminary study, with four patients and five sessions. Thus, we cannot predict the asymptotic performance of our schemes when working with many subjects and sessions. Some works applying a multi-session BMI intervention showed that after a certain number of sessions, recalibration of the decoder did not improve the decoding performance [8], [14]. Regarding the applicability of our results to different typologies of patients, we consider that our findings may also apply for stroke or other neurological injuries. However, further research has to be conducted to validate if the generalization across sessions or subjects is affected differently for other neurological populations. In addition, factors such as the classification algorithm chosen and the feature extraction process may play a role in the session-to-session functioning of the BMI systems, and hence, the study of their impact can be an interesting pursuit for future research.

The results presented in this paper are relevant as there is an increasing interest to include novel technologies in neurorehabilitation. The importance of recalibration for rehabilitative systems has also been evaluated for other physiological signals, such as the electromyography [25]. Although there is still no evidence demonstrating that higher BMI accuracy provides better rehabilitative outcomes in BMI-triggered therapies, it is reasonable to think that increasing the amount of times that the patient is positively rewarded may increase or accelerate the therapeutic effects. However, it is important to note that BMI accuracy is not the only factor affecting this rehabilitative process, as it has been demonstrated that the temporal precision is also of great importance in this respect [26].

ACKNOWLEDGMENT

This work has been supported by Spanish projects HYPER-CSD2009-00067, DPI2011-25892, and DGA-FSE (grupo T04).

REFERENCES

[1] H. Krebs, N. Hogan, B. Volpe, M. Aisen, L. Edelstein, and C. Diels, "Overview of clinical trials with mit-manus: a robot-aided neuro-rehabilitation facility," *Technology and Health Care*, vol. 7, no. 6, pp. 419–423, 1999.

[2] P. H. Peckham, M. W. Keith, K. L. Kilgore, J. H. Grill, K. S. Wuolle, G. B. Thrope, P. Gorman, J. Hobby, M. Mulcahey, S. Carroll *et al.*, "Efficacy of an implanted neuroprosthesis for restoring hand grasp in tetraplegia: a multicenter study," *Archives of physical medicine and rehabilitation*, vol. 82, no. 10, pp. 1380–1388, 2001.

[3] G. A. Prieto, R. Cano-de-la Cuerda, E. López-Larraz, J. Metrot, M. Molinari, and L. E. van Dokkum, "Emerging perspectives in stroke rehabilitation," in *Emerging Therapies in Neurorehabilitation*. Springer Berlin Heidelberg, 2014, pp. 3–21.

[4] J. J. Daly and J. R. Wolpaw, "Brain–computer interfaces in neurological rehabilitation," *The Lancet Neurology*, vol. 7, no. 11, pp. 1032–1043, 2008.

[5] E. López-Larraz, J. M. Antelis, L. Montesano, A. Gil-Agudo, and J. Minguez, "Continuous decoding of motor attempt and motor imagery from EEG activity in spinal cord injury patients," in *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012.

[6] J. M. Antelis, L. Montesano, A. Ramos-Murguialday, N. Birbaumer, and J. Minguez, "Continuous decoding of intention to move from contralesional hemisphere brain oscillations in severely affected chronic stroke patients," in *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012.

[7] E. López-Larraz, L. Montesano, A. Gil-Agudo, and J. Minguez, "Continuous Decoding of Movement Intention of Upper Limb Self-Initiated Analytic Movements from Pre-Movement EEG Correlates," *J Neuroeng Rehabil*, vol. 11, no. 153, 2014.

[8] 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*, vol. 59, no. 2, pp. 133–42, 2013.

[9] A. Ramos-Murguialday, D. Broetz, M. Rea, L. Läer, Ö. Yilmaz, F. L. Brasil, G. Liberati, M. R. Curado, E. Garcia-Cossio *et al.*, "Brain–machine interface in chronic stroke rehabilitation: a controlled study," *Annals of neurology*, vol. 74, no. 1, pp. 100–108, 2013.

[10] Y. Kasashima-Shindo, T. Fujiwara, J. Ushiba, Y. Matsushika, D. Kamatani, M. Oto, T. Ono, A. Nishimoto, K. Shindo, M. Kawakami, T. Tsuji, and M. Liu, "Brain Computer Interface Training Combined With Transcranial Direct Current Stimulation in Patients With Chronic Severe Hemiparesis: Proof of Concept Study," *Journal of Rehabilitation Medicine*, no. 15, 2015.

[11] B. Blankertz, F. Popescu, M. Krauledat, S. Fazli, M. Tangermann, K. Müller, R. Grinter, T. Rodden, P. Aoki, E. Cutrell *et al.*, "Challenges for brain-computer interface research for human-computer interaction applications," in *ACM CHI Workshop on Brain-Computer Interfaces for HCI and Games*, 2008.

[12] M. Takahashi, K. Takeda, Y. Otaka, R. Osu, T. Hanakawa, M. Gouko, and K. Ito, "Event related desynchronization-modulated functional electrical stimulation system for stroke rehabilitation: A feasibility study," *J Neuroeng Rehabil*, vol. 9, no. 56, 2012.

[13] K. Shindo, K. Kawashima, J. Ushiba, N. Ota, M. Ito, T. Ota, A. Kimura, and M. Liu, "Effects of neurofeedback training with an electroencephalogram-based brain-computer interface for hand paralysis in patients with chronic stroke: A preliminary case series study," *J Rehabil Med*, vol. 43, pp. 951–57, 2011.

[14] 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," in *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012.

[15] E. López-Larraz, I. Iturrate, L. Montesano, and J. Minguez, "Real-time recognition of feedback error-related potentials during a time-estimation task," in *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2010.

[16] 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," in *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011.

[17] E. López-Larraz, I. Iturrate, C. Escolano, I. Garcia, L. Montesano, and J. Minguez, "Single-trial classification of feedback potentials within neurofeedback training with an EEG brain-computer interface," in *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011.

[18] F. Lotte and C. Guan, "Learning from other subjects helps reducing brain-computer interface calibration time," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 2010, pp. 614–617.

[19] M. Krauledat, M. Tangermann, B. Blankertz, and K.-R. Müller, "Towards zero training for brain-computer interfacing," *PloS one*, vol. 3, no. 8, p. e2967, 2008.

[20] G. Müller-Putz, D. Zimmermann, B. Graimann, K. Nestinger, G. Korisek, and G. Pfurtscheller, "Event-related beta EEG-changes during passive and attempted foot movements in paraplegic patients," *Brain Research*, vol. 1137, pp. 84 – 91, 2007.

[21] C. Tangwiriyasakul, R. Verhagen, W. L. C. Rutten, and M. J. A. M. van Putten, "Temporal evolution of event-related desynchronization in acute stroke: a pilot study." *Clinical neurophysiology*, vol. 125, no. 6, pp. 1112–20, 2014.

[22] G. Pfurtscheller and F. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842 – 1857, 1999.

[23] H. Shibasaki and M. Hallett, "What is the Bereitschaftspotential?" *Clinical neurophysiology*, vol. 117, no. 11, pp. 2341–2356, 2006.

[24] L. Clemmensen, T. Hastie, D. Witten, and B. Ersbøll, "Sparse Discriminant Analysis," *Technometrics*, vol. 53, no. 4, pp. 406–413, 2011.

[25] A. Sarasola-Sanz, N. Irastorza-Landa, F. Shiman, E. López-Larraz, M. Spüler, N. Birbaumer, and A. Ramos-Murguialday, "EMG-based multi-joint kinematics decoding for robot-aided rehabilitation therapies," in *14th Int Conf on Rehabilitation Robotics*, 2015.

[26] N. Mrachacz-Kersting, S. R. Kristensen, I. K. Niazi, and D. Farina, "Precise temporal association between cortical potentials evoked by motor imagination and afference induces cortical plasticity." *The Journal of physiology*, vol. 590, no. Pt 7, pp. 1669–82, Apr. 2012.