View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Infoscience - École polytechnique fédérale de Lausanne

Long-term stable control of motor-imagery BCI by a locked-in user through adaptive assistance

Sareh Saeedi, Ricardo Chavarriaga, José del R. Millán

***Abstract*—Performance variation is one of the main challenges that BCIs are confronted with, when being used over extended periods of time. Shared control techniques could partially cope with such a problem. In this paper, we propose a taxonomy of shared control approaches used for BCIs and we review some of the recent studies at the light of these approaches. We posit that the level of assistance provided to the BCI user should be adjusted in real time in order to enhance BCI reliability over time. This approach has not been extensively studied in the recent literature on BCIs.**

In addition, we investigate the effectiveness of providing online adaptive assistance in a motor-imagery BCI for a tetraplegic end-user with an incomplete locked-in syndrome in a longitudinal study lasting 11 months. First, we report a reliable estimation of the BCI performance (in terms of command delivery time) using only a window of 1 s in the beginning of trials (AUC$\approx$ 0.8). Second, we demonstrate how adaptive shared control can exploit the output of the performance estimator to adjust online the level of assistance in a BCI game by regulating its speed. In particular, online adaptive assistance was superior to a fixed condition in terms of success rate ($p < 0.01$). Remarkably, the results exhibited a stable performance over several months without recalibration of the BCI classifier or the performance estimator.

***Index Terms*—Brain-computer interface, Motor imagery, Adaptive shared control, Performance estimation**

I. INTRODUCTION

Despite great progress in brain-computer interfaces (BCIs), both with able-bodied subjects and end-users, they are still hindered by uncertainty in the interpretation of brain signals. In this regard, one of the main issues is that BCI performance is inconsistent across and within subjects and fluctuates greatly over time [1]. Importantly, this issue is even more critical for motor-restricted end-users [2]–[9]. Therefore, in order for BCIs to be used reliably for extended periods of time, there is a need for overcoming such variations.

One way to tackle BCI performance variations is through the use of *shared control* approaches [10], [11]. These techniques are known to increase reliability by allowing both the human intentions (i.e., as decoded by BCI) and the controlled device to contribute in achieving the intended task. The use of shared control for BCI has shown to improve task performance and to reduce the user's workload [7], [9], [10], [12]–[15]. Nevertheless, shared control approaches for BCI typically rely on predefined settings, giving a fixed level of assistance to the user. Therefore, they cannot effectively mitigate variations in the users' performance over extended periods of time. In this work, we give an overview of the shared control approaches that have been implemented for BCIs and we propose to follow an approach that has not been extensively studied in the field. We hold that the system should be able to dynamically adapt to the user's condition and needs in order to successfully address performance variations. Such a kind of adaptive shared control approach should not only incorporate the state of the controlled device and its surrounding (i.e., *external context*), but should also consider the user *internal context*, such as cognitive states and brain signal reliability. In the following, we evaluate the effectiveness of such an adaptive shared control approach for an end-user with locked-in syndrome over a period of 11 months.

Related to the internal context, several studies have addressed performance variation in sensorimotor rhythm (SMR)-based BCIs [16]. These studies focus on inter-subject variability from an anatomical [17], [18], physiological [19]–[23], or psychological [24], [25] perspective or, in a lesser extent, on a combination of them [26]. In contrast, the intra-subject performance variation is less often investigated. Studies on this issue are dedicated to psychological aspects [27] or neurophysiological markers [21]–[23], [28]–[30]. In the latter case, efforts have been made for identifying neural correlates for performance variation extracted from EEG oscillations in different frequency bands and brain regions. For instance, the certainty of motor imagery (MI) classification was reported to be correlated with the power of gamma band oscillations [21], [22]. On the same line, it was suggested that a weighted combination of theta, alpha, and beta oscillations prior to the beginning of a trial is correlated with the MI classification performance [30]. Another study reported a positive correlation between the pre-cue SMR activity in subject-specific electrodes and frequency bands and the single-sample classification accuracy [28]. These studies provide important information on possible correlates of the MI performance. However, they do not make an estimation of the performance on a single-trial basis. Therefore, it is not possible to exploit this information online in order to change the interactions of the system based on the user's needs.

In this paper we investigate the use of internal context for BCI. In particular, we assess the feasibility of predicting MI-BCI performance (in terms of the command delivery time, CDT) on a trial-by-trial basis for a tetraplegic end-user with an incomplete locked-in syndrome. Following shared control principles, online adaptive assistance was provided based on the estimated CDT by modulating the speed of a BCI game.

S. Saeedi, R. Chavarriaga, and J. d. R. Millán are with Defitech Chair in Brain-Machine Interface (CNBI), Center for Neuroprosthetics, School of Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Geneva, Switzerland. e-mail: (sareh.saeedi, jose.millan)@epfl.ch

Manuscript received March 24, 2016; revised September 6, 2016.

Results show that our system yielded stable online performance across more than 30 sessions without re-calibration of either the MI-BCI classifier or the performance estimator despite long breaks in BCI usage due to the end-user's health condition. Furthermore, as it was the case in a previous study with able-bodied subjects [15], adaptive assistance led to higher success rates as compared to a fixed condition of the BCI game.

This paper is organized as follows, Section II discusses the state-of-the-art and introduces a taxonomy of the different approaches of shared control used in BCI applications. Section III-B details the experimental protocol in this study, followed by the obtained results (Section IV). Finally, Section V discusses the presented findings and future work.

II. Shared control approaches for BCI

For a reliable operation of assistive devices controlled by BCI, some degree of assistance (shared control) is required to overcome issues like low information throughput and BCI performance variations. Employing shared control techniques for BCIs have been shown to result in better performance, higher speed and safety, and lower perceived workload [9]. To provide shared control for BCI, different approaches have been implemented that can be summarized in mainly three different categories (Figure 1): *(i) contextual fusion*, where the internal and the external context contribute directly to the final control command, *(ii) contextual gating*, in which an initial command can be confirmed or refused by an internal or an external context, and *(iii) contextual regulation*, where an internal or an external context can refine the controller.

A. Contextual fusion

In this approach, the final control command is derived from a direct contribution of both the internal and the external context, Figure 1(a). In order to fuse the internal and external context in these systems, various techniques can be applied. All the behaviors (like obstacle avoidance, or goal-directed behaviors) assumed for a task should be merged together according to a specific rule. Some of the rules proposed for this purpose are competitive methods, weighted sums of different behaviors, and probabilistic reasoning [31].

This approach has been used to control a wheelchair or a telepresence robot [9], [11], [32]. In these applications, the user generates the steering commands (turning to the right or left) through the execution of two distinct motor-imagery tasks. This constitutes the internal context in this approach, which represents the user's intents. The device is, on the other hand, responsible for low level commands, including obstacle avoidance through the use of its sensors. In this way, the external context is provided for the shared control approach.

In the two mentioned applications, the information about the environment is used as the external context. However, any other source of information, which is not internal to the user, but could help them to accomplish the task can also serve as the external cue. For instance, a language model has been used in a hybrid-BCI text-entry system which implements the same two-class MI BCI as the one mentioned for the wheelchair and the robot [8]. Based on the language model, the characters are shown to the subject in a way to ensure the minimum average number of commands required to select each character. Therefore, the internal context for this shared control system consists of the users' decoded intention from the electroencephalogram (EEG) while the language model provides the external context.

B. Contextual gating

In this approach, the initial command constitutes a control command suggested to the user, which should be confirmed by the internal and the external context in order to be executed (Figure 1(b)). The initial command in this approach can be obtained from either a fusion or regulation framework (see next section).

Following the gating approach for shared control, a semi-autonomous navigation strategy for a mobile robot has been introduced where the proposed actions by the robot may be denied or approved via error-related EEG potentials (ErrP) [33]. In this study, the robotic system determines the possible decision points (e.g. crossings) by analyzing the environment via its embedded laser scanner. Then, the robotic system proposes the most probable action to the user by means of a visual feedback. This serves as the initial command for the system. Then, the user may confirm or refuse the proposed action via a brain channel, which makes the internal context. This approach is suitable for cases where either two actions have the same probability to be executed according to the robot controller or there is a conflict between the proposed action and the user's intention.

Following the same line, shared control principles have been used to investigate the feasibility of restoration of hand, finger and elbow function for a user with high-level spinal-cord injury (SCI) [7]. In this study, a hybrid neuroprosthesis consisting of functional electrical stimulation (FES) and a semi-active orthosis was used. A MI-based BCI switch was implemented to switch between hand and elbow control, while employing a gating mechanism which allowed for eliciting a BCI switch only when a shoulder joystick had not been moved in the previous seconds. This shared control scheme has shown to provide a reliable BCI control, even though only moderate BCI performance was achieved after extensive training, which might be inevitable especially for motor-disabled end-users.

C. Contextual regulation

In this framework, the level of assistance (shared control) can be regulated by means of an internal or an external context (Figure 1(c)). That is, the parameters of the controller can be adjusted while the system is being used. The initial command can be obtained by direct decoding of other brain signals, or even from the outcome of another shared control technique (i.e., contextual fusion or gating).

A semi-autonomous navigation system has been introduced following such a regulation approach [34]. Similar to the system described in the previous section, this system also involves proposing actions to the user who can confirm or refuse them. The difference, however, is that the user's habits are learned

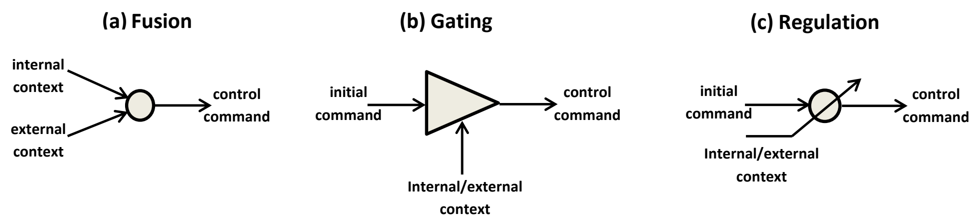

Fig. 1. Three different approaches for applying shared control principles for BCI systems: (a) Fusion: the internal and the external context contribute directly to the final control command. (b) Gating: an initial command can be confirmed or refused by an internal or an external context. (c) Regulation: an internal or an external context can refine the controller.

when navigating in a known environment so as to anticipate the next desired destination ahead of time. To this end, a dynamic Bayesian network (DBN) has been implemented to track the user's intended actions or goal destinations. This allows for adjusting to the user's instantaneous needs and in turn, reducing the user's workload.

A different approach has been taken in [35], which is based on human monitoring of the performance of an external autonomous device. In this study, ErrP was found in the EEG while the users received erroneous feedback corresponding to the cursor direction of movement (i.e., opposite to the target location). Upon identification of errors, learning of the optimal behavior could be achieved by decreasing the likelihood of repeating such decisions in the same context. In this way, the user provides reinforcement signals that can be used by the system to improve the overall performance.

Following the contextual regulation framework, we have introduced an approach for providing adaptive assistance to the user based on an assessment of their performance in executing mental commands in a MI task. This approach will be detailed in Section III-B.

III. Methods

A. Participant

The participant is a 53 year-old man with incomplete locked-in syndrome (LIS) after hemorrhagic brainstem stroke (4 years prior to the start of the experiment). He has been right-handed before the stroke occurred. Breathing is facilitated for him through tracheotomy. He is cognitively intact and his residual abilities include minor (and very slow) index finger movements (flexor muscle) of the left hand, eye blinks, eyebrow movements (corrugator supercilii muscles) and very limited neck movement. He wears glasses with the right glass lens blindfolded to allow focusing gaze, as the oculomotor muscles are completely paralyzed. Therefore, the use of eye tracker is not possible for him. Verbal communication is achieved through a custom-made solution based on lexicographic search assisted by a caregiver, in which the alphabet is split in four groups and each group consists of two rows of letters. He can select the letters with yes/no responses to a caregiver by means of the eye blinks/eyebrow movements, respectively.

B. Experimental protocol

The experiments were conducted in our laboratory facilities over a period of 11 months (c.f., Figure 2). Sessions, defined as a set of experimental tasks performed on a single day, were held regularly twice a week. Over the experiment duration, a significant number of sessions were skipped mainly due to illness or unavailability of the end-user. The gap between two consecutive sessions varied from a few days to 2 months. One recording session typically consisted of 3 to 4 runs of a two-class MI-BCI, which lasted less than 1 hour including the preparation time. Each run , referred to as a part of the experiment which is conducted in a time interval with no interruption, included 15 trials per class. The subject had previous experience with the MI protocol described in [5]. He was initially instructed to imagine the kinesthetic movement of the hands and to find the mental strategy that results in the best performance for him over the MI online recordings. The mental tasks he performed were the imagination of middle and third fingers of right hand for the right command and the thumb of left hand for the left command.

The experimental protocol consisted in a MI-BCI game depicted in Figure 3 [15]. In the online experiments, the subject controls the platform performing the two selected mental tasks (right/left hand MI) so that the parachutist lands on it. In fact, the MI classifier output is translated into the movement of the platform at each time point. The platform continues to move until the classifier output surpasses a subject-specific threshold, at which point the corresponding BCI command is 'delivered' and the platform reaches the left-most or right-most edge of the screen. If the platform reaches the target before the parachutist lands, it changes color to green and otherwise to red. In the default setup, the parachutist lands at second 10. As described

Fig. 2. Experimental setup showing the end-user, the EEG acquisition device and the computer screen showing the protocol feedback. In the experiment reported here, the horizontal cursor corresponding to the decoded MI task was replaced by the parachuting BCI game described in Figure 3.

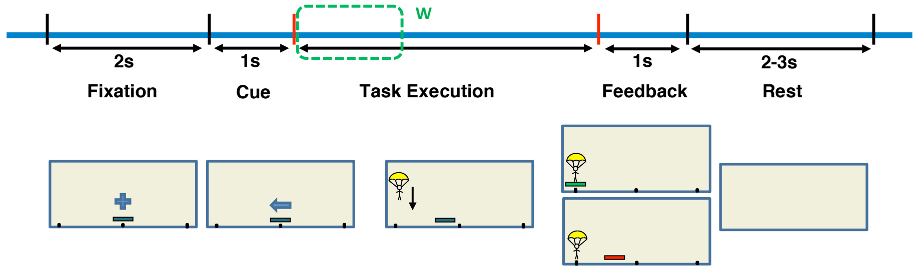

Fig. 3. The MI BCI game. The user has to bring the platform (the rectangle at the bottom of the screen) to the correct side for the parachutist to land safely. In the online experiments, the platform moves contineously to the left or to the right depending on the MI classifier output. At the beginning of each trial, the parachutists starts to vertically fall from the top of the screen (either on the left or the right edge). The parachutist lands at second 10 as default, but it can have different speeds depending on the level of assistance. If the platform reaches the target on time, it changes color to green and otherwise to red. See text for more details [15].

in Section III-E, in some phases of the experiment, this speed was regulated depending on the expected command delivery time of the current command sent by the user. Parachutist speed regulation provides the proper level of assistance to the user at a single-trial level. In the offline experiment, the parachutist lands on the platform which automatically moves towards the target within 4s without being controlled by the subject.

The choice of CDT as a performance metric is motivated by the nature of the MI-BCI system employed in this study (c.f. Section III-C), which integrates the classifier output over time in order to increase the reliability of decisions. A command is delivered as soon as the integrated output reaches a predefined threshold [5]. Therefore, BCI commands are delivered at different speeds, and CDT reflects how well the user can sustain MI to generate discriminant sensorimotor rhythms. Section III-D describes the performance estimator.

The longitudinal evaluation was carried out in 5 phases as described below (c.f., Table I).

Phase 1 involved training with the MI-BCI game in order to obtain a MI decoder with a rather stable performance. After an initial offline session, 9 online sessions were performed. Each session comprised 3 or 4 runs. During this phase, the MI classifier was recalibrated whenever the performance dropped (i.e., having a performance around 50% or biased performance towards one class). Since in previous experiments this subject had long command delivery times (CDT up to 45 s), we set a timeout of 10 s to avoid frustration, as suggested by [5]. Hereafter, we refer to this condition (i.e., when the timeout is 10 s) as the *normal MI*. After this phase, the parameters of the MI decoder were fixed and no further modification was done to them during the rest of the experiment (Phases 2-5).

In **Phase 2**, six sessions of online experiments were performed. As before, each session comprised 3 to 4 runs. The data of these sessions was used to build a performance estimator to differentiate between short and long CDT based on the initial samples of each trial (c.f., Section III-D). This performance estimator was then used without any further adjustment to provide online adaptive assistance during Phases 3 and Phase 5. That means, there was no re-calibration whatsoever to any decoder in the system from this moment on.

In **Phase 3**, we evaluated the use of online adaptive assistance. Three conditions were considered: *normal MI* (timeout equal to 10s), *adaptive assistance* (where the timeout is regulated based on an estimation of the CDT), and *fixed timeout* (timeout equal to 3s, see section III-E). All the sessions started with one run of the normal MI condition, followed by one run of each of the other two conditions. In 5 sessions out of 11 (i.e., S16-19 and S24), the fixed timeout was performed before the adaptive assistance condition, and the order of the two was reversed in the remaining 6 sessions. Typically, runs consisted of 20 trials per mental task[1]. In this phase, there were several gaps between sessions due to end-user's health conditions (see Table I), which finally led to an interruption of 2 months.

Phase 4 started after this break. We performed four online sessions of *normal MI* (i.e., without assistance) to allow the end-user to regain his skills in MI BCI. During this and the following phase, the length of each run was reduced to 15 trials per mental task in order to avoid excessive fatigue.

Phase 5 was also devoted to investigating the effectiveness of providing online adaptive assistance in order to assess the performance for a longer period of time. We made some changes in Phase 5 to have a more reasonable evaluation of the performance in different conditions. First, we preferred not to conduct the fixed timeout condition and to estimate it from the normal MI condition in order to reduce the fatigue resulting from long recordings. Second, all the 10 sessions of Phase 5 consisted of 4 runs, where the first 2 runs were always normal MI, followed by adaptive assistance condition. The order of the other two runs was adaptive assistance first, followed by normal MI in the first 5 sessions and vice versa in the remaining ones. These changes allowed for having more number of runs per condition in the same amount of time.

C. Motor Imagery BCI

MI training was performed similar to what has been explained in [5]. The difference, however, was the game, which represented a BCI paradigm with time constraints. Also, this game provided a more engaging environment for the users [15]. EEG was recorded using 16 electrodes over the sensorimotor cortex at 512 Hz and band-pass filtered between 0.1 Hz and 100 Hz. Laplacian spatial filtering was applied to the signal. A MI classifier was trained to discriminate between the

[1]Due to the user's condition during sessions 24 and 25, he only did 10 trials per mental task.

		Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
Number of sessions		1+9	6	11	4	10
Number of runs per session	Normal MI (timeout: 10 s)	3-4	3-4	1	4	2
	Fixed timeout (timeout: 3 s)	—	—	1	—	—
	Adaptive assistance (timeout: 3/10 s)	—	—	1	—	2
Number of interruptions		3	2	6	1	2
Duration of interruptions		4 weeks	4 weeks	6 weeks	1 week	2 weeks
Average success rate		0.56 ± 0.18	0.63 ± 0.04	0.67 ± 0.10	0.66 ± 0.07	0.69 ± 0.07
MDT (s) , IQR (s)		4.3 , 7.8	5.4 , 7.6	3.9 , 7.0	3.0 , 5.0	3.2 , 6.1

TABLE I
FIVE PHASES OF THE EXPERIMENT. IN THE LAST ROW, MDT AND IQR REFER TO MEDIAN DELIVERY TIME AND INTERQUARTILE RANGE, RESPECTIVELY.

two mental tasks in a session with offline recordings (Phase 1, Table I). The procedure for building this classifier is described in more detail in [5]. First, power spectral densities (PSDs) were extracted from the EEG signal in the range of 4-48 Hz with the resolution of 2 Hz. Given the number of channels (16) and the number of frequency components (23), each EEG sample comprises 368 features. Then, a subset of the features were selected based on their relevance to the mental tasks using canonical variate analysis (CVA) [36] and the neurophysiological evidence on the cortical areas/frequency bands contributing to a specific mental task. Out of all the PSD features, 5 to 7 were selected in the alpha and beta band, which were then used by a Gaussian classifier with four prototypes per class to infer the probability of each sample belonging to a class. In the online sessions, the classifier output is aggregated over time using a leaky integrator in order to increase the reliability of decisions [5]. The feedback (i.e. platform movements) are ruled then by the integrated evidence. Consequently, the time it takes to move the platform to the target location (i.e., command delivery time) will vary depending on the output of the MI classifier. Therefore, the delivery time can be used as a measure of the system performance.

D. Single-trial prediction of short vs. long commands

The data recorded in Phase 2 (670 trials) was used to build a decoder to predict whether the current trial will have a short or long delivery time. During this phase the subject had a mean success rate of 63% (Table I), while only 8% of the commands were erroneous and 29% were timeout (i.e., the classification accuracy did not reach the threshold within 10 s). Figure 4(a) illustrates the distribution of the CDT in Phase 2 considering both correct (hit) and timeout commands. In order to better visualize the CDT distribution for correct commands, Figure 4(b) only considers these commands and discards the timeout ones.

Figure 4(c) shows the characteristics of different types of commands (i.e., short, long, and timeout) in terms of the average single-sample classification accuracy over the trials. In this case, we considered accurate commands delivered in less than 3 s as being short. As shown in the figure, the average single-sample accuracy is the highest for the short commands and lowest for timeout commands, while long commands have an accuracy in between. This shows the link between CDT and the MI classification accuracy, supporting the choice of using it as a measure of the reliability of the BCI channel.

In order to build the performance estimator, the erroneous commands were excluded. Given the considerable number of timeout commands, two cases were compared: using only hit commands for training the classifier or including the timeout ones as well. The classification performance was compared for these two cases using a separate set of data.[2]

As explained in [15], for the classification of short vs. long commands, different percentiles of the CDT were considered to separate the trials into the relevant groups. The performance estimator uses the same PSD features as the MI classifier. We hypothesize that the evolution of these features over time might reflect the quality of an ongoing command. Therefore, a window of 1 s in the beginning of the trials (the green window in Figure 3) was used to make a prediction about the CDT (long vs. short). A separate classifier was built per mental task (right-C1 and left-C2) as the characteristics of the two classes might be different.

For estimating the trial performance given the feature vectors ($\mathbf{x}$) extracted from EEG within W, we define a distance measure for class i and Gaussian prototype j as:

$$Dist_{ij} = \frac{1}{N_w} \sum_{t=1}^{N_w} \sum_{k=1}^{N_f} \frac{(x_{tk} - \mu_{ijk})^2}{\sum_{ik}}, \tag{1}$$

where $\mathbf{x}$ is the feature vector with N_f selected features (i.e. EEG channels and frequency bands), μ_{ij} is the center of the j^{th} prototype of class i, Σ_{ik} is the variance of feature k for class i, and N_w is the number of feature vectors within the

[2]To evaluate the classification performance, a testing set of data was selected which was not included in the training set. This dataset consisted of 6 runs recorded in 4 different sessions (between Phase 2 and Phase 3). These recordings were not included in Phases 1-5 described here, as the number of runs in a single session was less than 3 due to technical problems or the participant's fatigue. The average success rate and the CDT distribution was comparable to the training set (i.e., the data in Phase 2). In this dataset, the percentage of hit, miss, and timeout commands was 64%, 9%, and 26%, respectively.

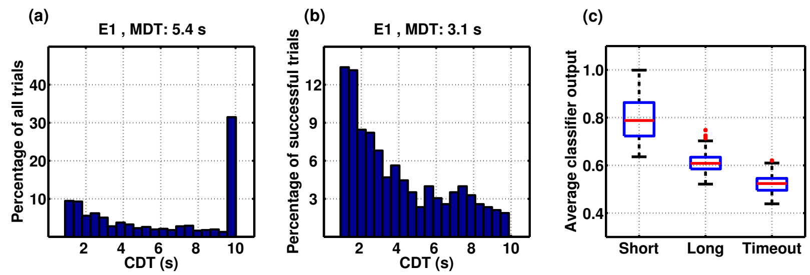

Fig. 4. Distribution of CDT for the end-user (E1) when (a) all the correct (hit) and timeout commands are considered, and (b) when only hit commands are considered. (c) The average classifier sample accuracy for short, long, and timeout commands.

window. Given that we have N_p (equal to 4) prototypes per class, the feature vector f_{sl} for estimating the performance is composed of the distance of sample x_t to every prototype ($f_{sl} = Dist_{ij}, i = 1 : N_c, j = 1 : N_p$). Since we consider that the distance to some of the prototypes might contribute more to the discrimination of the short and long trials, a subset of them were selected using CVA. Finally, a linear discriminant analysis (LDA) classifier per mental task was built to differentiate between short and long commands. The classification performance was evaluated using 10-fold cross validation. We also compared different thresholds for differentiating long and short commands. For this, we considered different percentiles of the distribution command delivery time (from 35 to 65, with a step size of 5). Then, a threshold t_{sl} (corresponding to a percentile of the CDT) was chosen, achieving a reasonable classification performance (Area under the ROC curve, AUC $\approx>$ 0.8 for both classes).

E. Online adaptive assistance

Based on the performance achieved for classification of short vs. long commands on the data of Phase 2, $t_{sl} = 3\ s$ was selected as the threshold between long and short trials. Therefore, in the fixed timeout condition, the timeout was set to 3 s. The two conditions (i.e., fixed timeout and adaptive assistance) were compared in Phase 3 and Phase 5. As the fixed timeout condition was not performed in Phase 5, the success rate (the ratio of correct commands over all commands) for this condition was estimated from the normal MI condition by assuming a timeout of 3 s. In order to evaluate the performance in each condition with respect to the chance level for classification, the online command accuracy (computed as the ratio of correct commands over the number of correct and wrong commands only, discarding timeout commands) has been considered. More details can be found in the appendix in Secion VI.

Also, it is essential to evaluate whether the proposed online adaptive assistance performs better than providing random assistance. The threshold $t_{sl} = 3\ s$ corresponds to the 35^{th} percentile of the distribution of the CDT (hit and timeout commands combined) in Phase 2. That is, 65% of the commands are longer than t_{sl} while the rest are successfully completed before. Therefore, we defined the random case as having a timeout of 3 s by default and randomly providing assistance (i.e., changing the timeout to 10 s) for 65% of the total commands. Then, we compared the number of hit commands per minute for random assistance and adaptive assistance cases. The random assistance case was simulated 20 times using the normal MI recordings in the corresponding sessions and the average number of hits per minutes was reported.

IV. Results

A. Event-related desynchronization (ERD) comparison for trials of different length

To characterize different types of command delivery in Phase 2 (i.e., short, long, and timeout), ERD was calculated as the ratio of EEG mu band (8-14 Hz) power in a window of 2 s at the end of each trial to the same window length before the cue appears (i.e., the fixation period in Figure 3). Figure 5 illustrates these differences in the the topographic maps of the ERD over the scalp for C1 (right hand) and C2 (left hand). Localized contralateral ERD is observed in the mu band for short commands of both classes, which is not the same for the long and timeout commands.

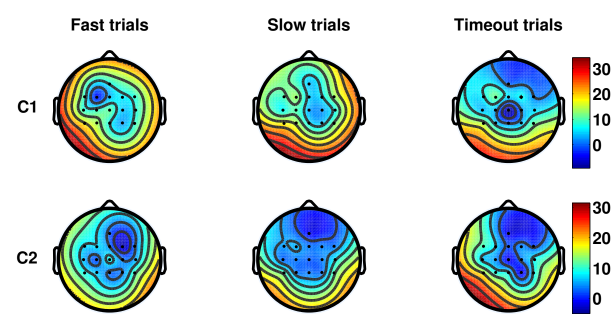

Fig. 5. Topography of the ERD (in percentage) in mu band (8-14 Hz), computed over a window of 2 s in the end of trials for short, long, and timeout commands. C1 and C2 correspond to right hand and left hand imagination of movement, respectively.

B. MI performance

The longitudinal nature of this study allows us to reliably evaluate fluctuations in MI performance (both in terms of success rate and command delivery time) over time. Figure 6 and Table I report the success rate and the error rate in the normal MI case over 40 sessions. The difference between the two is the rate of timeout commands. As depicted in the figure, during phase 1 –where the MI classifier had to be retrained

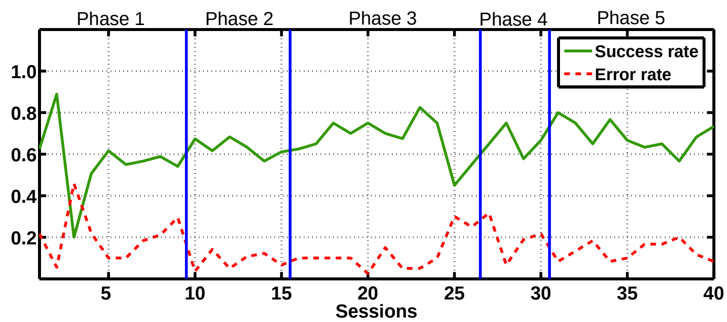

Fig. 6. Success rate and error rate in the normal MI condition over 40 sessions of the experiment.

several times– the success rate has a large variability (mean $\pm$ SD:0.56 ± 0.18).

Once the MI classifier is fixed at the beginning of Phase 2, the variability of the success and error rates are reduced compared to Phase 1. The only exception to this pattern appears at the end of phase 3 (session 25), where performance dropped. It is worth to mention that in sessions 24 and 25, only 10 trials per mental task have been performed due to the health condition of the user, what probably impaired a proper estimation of the actual performance given the low number of trials. As mentioned before, to compensate for this drop, 4 sessions of online MI (without adaptation) were performed allowing the subject to regain his previous performance (Phase 4). As the experiment proceeds to Phases 3-5, the success rate is higher with less variation (c.f., average performances reported in Table I). Furthermore, the error rate and the number of timeout commands are rather stable over time.

As the distribution of CDT is usually skewed, the median delivery time (MDT) and the interquartile range (IQR) are given in Table I for the normal MI condition in different phases. The MDT and the IQR are higher in Phase 1 compared to the other phases. Once the MI classifier is fixed (starting from Phase 2), they tend to decrease. That is, the speed of command delivery has been improved in Phases 3-5 with respect to Phases 1-2.

C. Single-trial prediction of short vs. long commands

Figure 7 illustrates the average AUC for short vs. long classification in Phase 2 (10 fold cross-validation). We report the classification performance considering different percentiles of CDT for separating the trials (ranging from 35th to the 65th percentile). For this classification, only hit commands were taken into account. Compared to the case with both hit and timeout commands, the case considering only hit commands resulted in a better performance on a dataset which was not included for classification (6 runs from 4 different sessions which were not included in the analysis of different phases, c.f. footnote 2). As depicted in Figure 7, a reasonable classification is achieved (AUC $\approx$ 0.8) for different percentiles. Classification performance drops faster for C1 after 50th percentile (around 3.5 s). In order to be consistent across the two classes, $t_{sl} = 3$ was selected which showed a reliable classification performance both for C1 and C2.

D. Online adaptive assistance

Figure 8 shows the success and error rates over the 11 sessions (S16 to S26) of Phase 3 for the adaptive assistance and the fixed timeout conditions. As seen in the figure, success rate in the adaptive assistance case is significantly higher than for the fixed timeout condition ($p < 0.01$; Wilcoxon rank sum test). The average success rate over the 11 sessions in this phase was 0.58 ± 0.12 for the adaptive assistance condition and 0.30 ± 0.06 for the fixed timeout condition. It is also worth noting that this figure shows an increasing trend in the success rate for the adaptive assistance condition (S16 to S24), while there is no such pattern for the fixed timeout condition. We also observe a performance drop in the adaptive assistance condition in S25 and S26. Remarkably, although there was a 1 month gap between sessions 23 and 24, the success rate in the adaptive assistance condition reaches 0.8 with a low error rate of 0.02.

Results during Phase 5 also yielded a significant improvement of the success rate when providing adaptive assistance ($p < 0.01$; Wilcoxon rank sum test). The average success rate over the 10 sessions in this phase was 0.60 ± 0.07 for adaptive assistance condition and 0.42 ± 0.08 for fixed timeout condition (Figure 9). When we compared the use of adaptive assistance with a system that provides assistance in a random manner we see that the former resulted in higher number of hit commands per minute in most of the sessions (Figure 10).

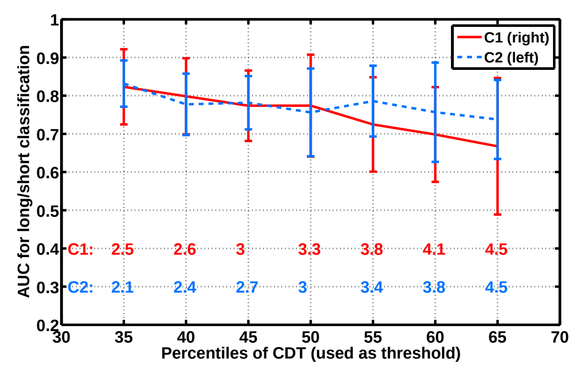

Fig. 7. Average AUC and standard deviation for short vs. long command classification (10-fold cross validation) for C1 (right, red) and C2 (left, blue), using different percentiles of CDT. At the bottom of the figure, the time threshold corresponding to each percentile is shown.

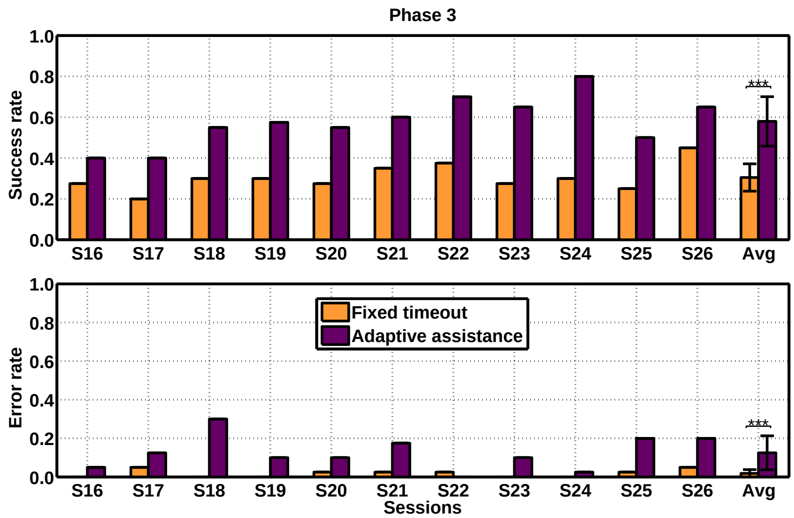

Fig. 8. Success rate and error rate in Phase 3: comparison between fixed timeout and adaptive assistance ($*\,*\,*\; p < 0.001$; Wilcoxon rank sum test).

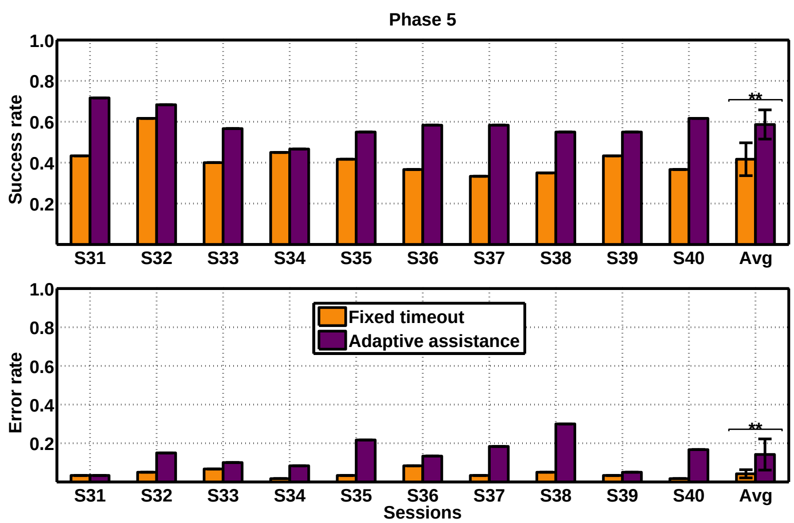

Fig. 9. Success rate and error rate in Phase 5: comparison between fixed timeout and adaptive assistance ($*\,*\;\; p < 0.01$; Wilcoxon rank sum test).

V. Discussion

In this study, we introduced, a taxonomy of shared control approaches employed for BCIs. We claim that these shared control systems are one or a combination of three main distinct frameworks: contextual fusion, contextual gating, and contextual regulation. We have reviewed the recent literature in the light of this taxonomy. Among the introduced frameworks, contextual fusion is the one most often employed in BCI applications. This approach is usually associated with predefined settings based on the task and the environment. However, to develop a reliable BCI which operates over extended periods of time, it must adapt to the users evolving needs. We argue that this adaptation should optimally be a function of both internal and external context. We suggest that either contextual gating or contextual regulation are more suitable to handle these requirements.

As an instantiation of contextual regulation shared control framework, we have investigated the feasibility of providing online adaptive assistance based on an estimation of performance for a locked-in syndrome end-user. With the insights from our previous study with able-bodied individuals [15], this longitudinal study allowed us to evaluate the effectiveness of such a method over an extended period of time.

This experiment was conducted in five different phases over a total of 40 sessions. The important characteristic of this experiment was the use of a fixed MI classifier for 31 sessions (from Phase 2 to Phase 5; a total duration of 9 months). It is noteworthy that even without recalibrating the classifier, the user could sustain a rather stable performance over time (Figure 6). The achieved success rate is in line with what has been reported for motor-restricted end-users in the literature [3], [7], [37], [38]. In particular, the average success rate (Table I) is similar to the one achieved by an individual with SCI (with the lesion at the level of C4 and the ability to control shoulder movements), where the classifier

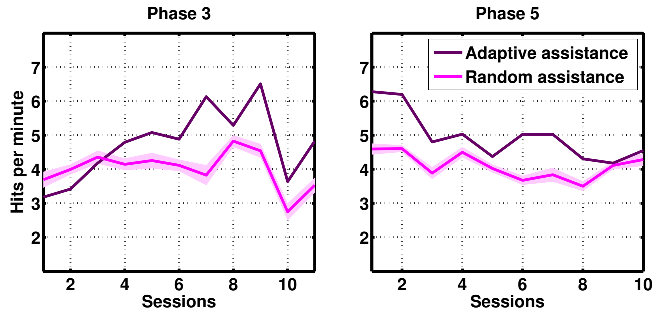

Fig. 10. Number of correct commands (hit) per minute: comparison between adaptive assistance and random assistance.

was re-calibrated at each session during the 11 months of experiment [7] . In addition, in a study on individuals with ALS, it has been reported that only 1 out of 7 tetraplegic participants could achieve such MI performance, while the 5 paraplegic participants could acquire a similar accuracy [3]. Therefore, it is notable that the end-user in this study has achieved an average performance similar to the paraplegic end-users. In fact, as presented in Table I, the performance showed an improvement both in terms of the success rate and the CDT from Phase 2 to Phase 5. It should be also noted that the successful commands tend to be delivered faster in the last phase of the experiment (as defined by the MDT and IQR in table I for different phases).

Importantly, the results on the classification of short vs. long commands with the end-user, replicate those previously obtained with able-bodied individuals [15]. The short and long commands could reliably be differentiated (AUC ≈ 0.8) using only a window of 1 s in the beginning of each trial. This gave us a measure of the user's performance, as the CDT is correlated with the average single-sample classification accuracy. That is, the short trials are the ones with the higher single-sample classification accuracy (Figure 4). Also, the pattern of activity in the brain differs for different command types i.e., short, long, and timeout. As depicted in Figure 5, a localized contralateral ERD is only observed for short trials when considering a window of 2 s in the end of commands, which is in line with the expected ERD patterns in MI for motor-restricted users [3].

The designed performance estimator was then used to provide online adaptive assistance based on the prediction of delivery time. That is, whenever a long trial was predicted, the system slowed down to give the user sufficient time. As for the healthy users, we compared two different conditions in Phases 3 and 5: adaptive assistance and fixed timeout. As illustrated in Figure 8, the success rate is significantly higher when providing adaptive assistance compared to using a fixed timeout. It should be noted that there were six interruptions within Phase 3 due to illness. Therefore, the performance degradation observed at the end of this phase could be explained as an effect of illness and medications. However, an increasing trend in the success rate for the adaptive assistance condition can be observed, which suggest a learning process.

After an interruption of 2 months, the user performed the normal MI condition in Phase 4, in order to restore the MI-BCI skill. In phase 5, normal MI and adaptive assistance were conducted. The fixed timeout was estimated from the normal MI condition. The comparison between the success rates in the adaptive assistance and the fixed timeout conditions reveals significant improvement of the former. This is in line with the results achieved for able-bodied individuals [15].

It has been suggested that the trial duration should be selected so as to obtain the highest bit rate [39]. However, this approach might be counterproductive, since it may increase the workload and stress levels for the user if the pace of the task is too fast. This is supported by the evaluation of the perceived workload using the NASA-TLX in able-bodied subjects [15][3]. Also, due to performance variations, a fixed setting like that may fail to be effective for a long time. Our results suggest that regulating the level of assistance based on the user's condition (i.e. internal context) is beneficial in this respect. Comparing the adaptive assistance with a random assistance clearly shows this benefit as illustrated in Figure 10. The random assistance here is based on the distribution of the CDT in the calibration phase which seems not to follow the inter-session changes.

Overall, the results suggest that providing adaptive assistance through the estimation of the CDT is particularly beneficial for this end-user. It should be noted that for this end-user, when no assistance was provided, the number of erroneous commands were rather low and stable, whereas timeout commands were more problematic. It remains to be studied whether this approach generalises to other users with severe motor disabilities and how other aspects of performance variation (e.g., accuracy) can be considered for subjects who show higher number of erroneous commands.

VI. Appendix

The online success rate reported in section IV-D is estimated considering the ratio of correct commands over all the commands (including wrong and timeout ones). However, in many applications (like a spelling task), timeout commands are considered as no command rather than wrong ones. Therefore, it is the online command accuracy (computed as the ratio of correct commands to the number of correct and wrong commands only, discarding timeout commands) that should be compared to the chance level for classification. To do so, we have computed the confidence intervals for $\alpha = 0.05$. Since the classification of the two-class motor imagery task follows a binomial distribution, we estimate the confidence intervals using Jeffreys' Beta distribution [40]. This method has been used to achieve a reasonable estimation even for a small number of trials (N), as it is adequate for $N > 10$. Approximation of the binomial distribution by a normal distribution (as implemented in [41]) is only valid if there are a certain number of trials ($N \times p \times (1-p) > 5$, where p is the prior) [40].

Figures 11 and 12 illustrate the online accuracy (ignoring timeout commands) as well as the chance level (red line) for Phase 3 and Phase 5, respectively. As depicted in these two figures, the online accuracy in both conditions is higher than random classification for a confidence of 0.95.

[3]The workload evaluation was not performed with the end-user since answering the questionnaire during each session will demand too much effort given his condition and communication capabilities.

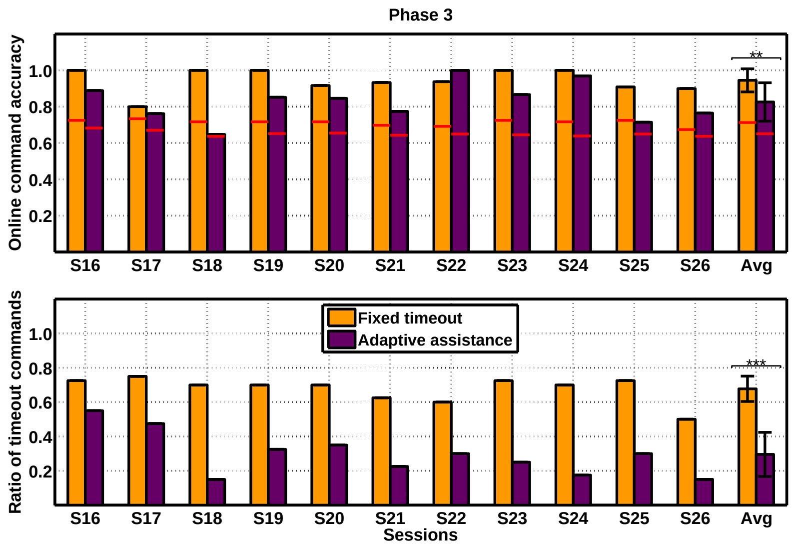

Fig. 11. Online command accuracy and ratio of timeout commands to all commands in Phase 3, comparing adaptive assistance and fixed timeout conditions ($** \ p < 0.01$, $*** \ p < 0.001$; Wilcoxon rank sum test). The red lines show the chance level for classification with the confidence interval of $\alpha = 0.05$.

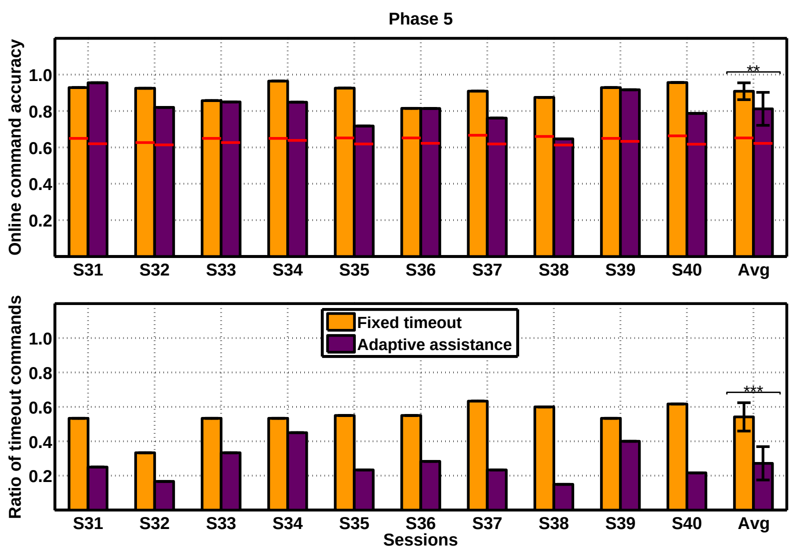

Fig. 12. Online command accuracy and ratio of timeout commands to all commands in Phase 5, comparing adaptive assistance and fixed timeout conditions ($** \ p < 0.01$, $*** \ p < 0.001$; Wilcoxon rank sum test). The red lines show the chance level for classification with the confidence interval of $\alpha = 0.05$.

Even though the online command accuracy is significantly higher in the Fixed timeout condition compared to Adaptive assistance, it should be noted that this is achieved at the expense of significantly higher number of timeouts ($p < 0.001$; Wilcoxon rank sum test) , and in consequence a lower success rate (c.f. Figures 8 and 9). In fact, in this case the pace of the task might be very fast for the user. This has been confirmed by evaluation of the perceived workload through NASA-TLX questionnaire in our previous work with healthy subjects [15].

In addition, Figure 13 illustrates an estimation of the chance level for classification in the normal MI condition over Phase 1 to Phase 5 during the 11 months of the experiment. This figure confirms that in almost all sessions, the online command accuracy (ratio of correct commands over the number of correct and wrong commands) is above the chance level for a confidence of 0.95. In sessions 24 and 25, only 20 commands were delivered, out of which around 20% were timeout. Therefore, this might not be sufficient to reliably compute the online command accuracy and it can explain the fact that the accuracy is lower than the chance level. In other sessions the number of commands were always more than 40.

ACKNOWLEDGMENT

This work is supported by Swiss National Center of Competence in Research (NCCR) Robotics and Hasler foundation. The authors would like to thank the end-user and his caregiver for their collaboration and participation in this study.

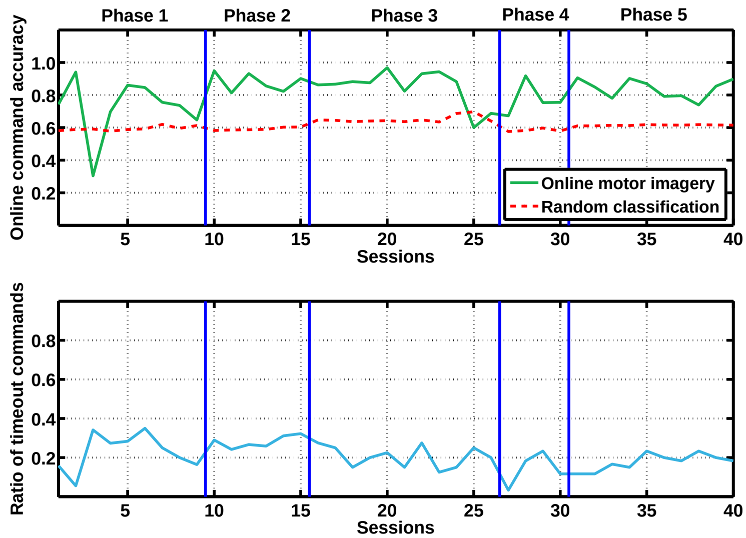

Fig. 13. Online command accuracy and ratio of timeout commands to all commands in the normal MI condition over Phases 1 to 5. The red lines show the chance level for classification with the confidence interval of $\alpha = 0.05$.

References

[1] B. Z. Allison and C. Neuper, "Could anyone use a BCI?" in *Brain-computer interfaces*, D. S. Tan and A. Nijholt, Eds. Springer, 2010, pp. 35–54.

[2] A. Kübler, N. Neumann, B. Wilhelm, T. Hinterberger, and N. Birbaumer, "Predictability of brain-computer communication," *Journal of Psychophysiology*, vol. 18, no. 2/3, pp. 121–129, 2004.

[3] G. Pfurtscheller, P. Linortner, R. Winkler, G. Korisek, and G. Müller-Putz, "Discrimination of motor imagery-induced EEG patterns in patients with complete spinal cord injury," *Computational Intelligence and Neuroscience*, vol. 2009, no. 1, 2009.

[4] K. Gourab and B. D. Schmit, "Changes in movement-related β-band EEG signals in human spinal cord injury," *Clinical Neurophysiology*, vol. 121, no. 12, pp. 2017–2023, 2010.

[5] R. Leeb, S. Perdikis, L. Tonin, A. Biasiucci, M. Tavella, M. Creatura, A. Molina, A. Al-Khodairy, T. Carlson, and J. d. R. Millán, "Transferring brain–computer interfaces beyond the laboratory: successful application control for motor-disabled users," *Artificial Intelligence in Medicine*, vol. 59, no. 2, pp. 121–132, 2013.

[6] M. Grosse-Wentrup and B. Schölkopf, "A review of performance variations in SMR-based brain- computer interfaces (BCIs)," in *Brain-Computer Interface Research*, C. Guger, B. Z. Allison, and G. E. Edlinger, Eds. Springer, 2013, pp. 39–51.

[7] 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. 133–142, 2013.

[8] S. Perdikis, R. Leeb, J. Williamson, A. Ramsay, M. Tavella, L. Desideri, E.-J. Hoogerwerf, A. Al-Khodairy, R. Murray-Smith, and J. d. R. Millán, "Clinical evaluation of BrainTree, a motor imagery hybrid BCI speller," *Journal of Neural Engineering*, vol. 11, no. 3, p. 036003, 2014.

[9] R. Leeb, L. Tonin, M. Rohm, L. Desideri, T. Carlson, and J. d. R. Millán, "Towards independence: A BCI telepresence robot for people with severe motor disabilities," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 969–982, 2015.

[10] L. Tonin, R. Leeb, M. Tavella, S. Perdikis, and J. d. R. Millán, "The role of shared control in BCI-based telepresence," in *2010 IEEE International Conference on Systems, Man and Cybernetics*, 2010, pp. 1462–1466.

[11] T. Carlson and J. d. R. Millán, "Brain-controlled wheelchairs: A robotic architecture," *IEEE Robotics and Automation Magazine*, vol. 20, pp. 65–73, 2013.

[12] J. d. R. Millán, F. Renkens, J. Mouriño, and W. Gerstner, "Noninvasive brain-actuated control of a mobile robot by human EEG," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1026–1033, 2004.

[13] F. Galán, M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, and J. d. R. Millán, "A brain-actuated wheelchair: asynchronous and non-invasive brain–computer interfaces for continuous control of robots," *Clinical Neurophysiology*, vol. 119, no. 9, pp. 2159–2169, 2008.

[14] I. Iturrate, R. Chavarriaga, L. Montesano, J. Minguez, and J. d. R. Millán, "Teaching brain-machine interfaces as an alternative paradigm to neuroprosthetics control," *Scientific Reports*, vol. 5, no. 13893, 2015.

[15] S. Saeedi, R. Chavarriaga, R. Leeb, and J. d. R. Millán, "Adaptive assistance for brain-computer interfaces by online prediction of command reliability," *Computational Intelligence Magazine, IEEE*, vol. 11, no. 1, pp. 32–39, 2016.

[16] M. Ahn and S. C. Jun, "Performance variation in motor imagery brain–computer interface: A brief review," *Journal of Neuroscience Methods*, vol. 243, pp. 103–110, 2015.

[17] S. Halder, D. Agorastos, R. Veit, E. M. Hammer, S. Lee, B. Varkuti, M. Bogdan, W. Rosenstiel, N. Birbaumer, and A. Kübler, "Neural mechanisms of brain–computer interface control," *NeuroImage*, vol. 55, no. 4, pp. 1779–1790, 2011.

[18] S. Halder, B. Varkuti, M. Bogdan, A. Kübler, W. Rosenstiel, R. Sitaram, and N. Birbaumer, "Prediction of brain-computer interface aptitude from individual brain structure," *Frontiers in Human Neuroscience*, vol. 7, no. 105, 2013.

[19] C. Sannelli, M. Braun, M. Tangermann, and K.-R. Müller, "Estimating noise and dimensionality in BCI data sets: Towards illiteracy comprehension," in *Proceedings of the 4th International BrainComputer Interface Workshop and Training Course 2008, Verlag der Technischen UniversitŁt Graz, Graz*, 2008, pp. 26–31.

[20] B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K.-R. Müller, G. Curio, and T. Dickhaus, "Neurophysiological predictor of SMR-based BCI performance," *NeuroImage*, vol. 51, no. 4, pp. 1303–1309, 2010.

[21] M. Grosse-Wentrup, B. Schölkopf, and J. Hill, "Causal influence of gamma oscillations on the sensorimotor rhythm," *NeuroImage*, vol. 56, no. 2, pp. 837–842, 2011.

[22] M. Ahn, S. Ahn, J. H. Hong, H. Cho, K. Kim, B. S. Kim, J. W. Chang, and S. C. Jun, "Gamma band activity associated with BCI performance: simultaneous MEG/EEG study," *Frontiers in Human Neuroscience*, vol. 7, no. 848, 2013.

[23] M. Ahn, H. Cho, S. Ahn, and S. C. Jun, "High theta and low alpha powers may be indicative of BCI-illiteracy in motor imagery," *PLoS ONE*, vol. 8, no. 11, 2013.

[24] E. M. Hammer, S. Halder, B. Blankertz, C. Sannelli, T. Dickhaus, S. Kleih, K.-R. Müller, and A. Kübler, "Psychological predictors of SMR-BCI performance," *Biological Psychology*, vol. 89, no. 1, pp. 80–86, 2012.

[25] A. Vuckovic and B. A. Osuagwu, "Using a motor imagery questionnaire to estimate the performance of a brain–computer interface based on object oriented motor imagery," *Clinical Neurophysiology*, vol. 124, no. 8, pp. 1586–1595, 2013.

[26] C. Jeunet, B. N'Kaoua, S. Subramanian, M. Hachet, and F. Lotte, "Predicting mental imagery-based BCI performance from personality, cognitive profile and neurophysiological patterns," *PloS One*, vol. 10, no. 12, p. e0143962, 2015.

[27] F. Nijboer, N. Birbaumer, and A. Kübler, "The influence of psychological state and motivation on brain–computer interface performance in patients with amyotrophic lateral sclerosis–a longitudinal study," *Frontiers in Neuroscience*, vol. 4, 2010.

[28] C. L. Maeder, C. Sannelli, S. Haufe, and B. Blankertz, "Pre-stimulus sensorimotor rhythms influence brain–computer interface classification performance," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 20, no. 5, pp. 653–662, 2012.

[29] M. Grosse-Wentrup and B. Schölkopf, "High gamma-power predicts performance in sensorimotor-rhythm brain–computer interfaces," *Journal of Neural Engineering*, vol. 9, no. 4, p. 046001, 2012.

[30] A. Bamdadian, C. Guan, K. K. Ang, and J. Xu, "The predictive role of pre-cue EEG rhythms on MI-based BCI classification performance," *Journal of Neuroscience Methods*, vol. 235, pp. 138–144, 2014.

[31] R. Leeb, H. Sagha, R. Chavarriaga, and J. d. R. Millán, "A hybrid brain–computer interface based on the fusion of electroencephalographic and electromyographic activities," *Journal of Neural Engineering*, vol. 8, no. 2, p. 025011, 2011.

[32] T. Carlson, R. Leeb, G. Monnard, A. Al-Khodairy, and J. d. R. Millán, "Driving a BCI wheelchair: A patient case study," in *Proceedings of TOBI Workshop lll*, 2012, pp. 59–60.

[33] X. Perrin, R. Chavarriaga, F. Colas, R. Siegwart, and J. d. R. Millán, "Brain-coupled interaction for semi-autonomous navigation of an assistive robot," *Robotics and Autonomous Systems*, vol. 58, no. 12, pp. 1246–1255, 2010.

[34] X. Perrin, F. Colas, C. Pradalier, R. Siegwart, R. Chavarriaga, and J. d. R. Millán, "Learning user habits for semi-autonomous navigation using low throughput interfaces," in *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, 2011, pp. 1–6.

[35] R. Chavarriaga and J. d. R. Millán, "Learning from EEG error-related potentials in noninvasive brain-computer interfaces," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 18, no. 4, pp. 381–388, 2010.

[36] F. Galán, P. W. Ferrez, F. Oliva, J. Guardia, and J. d. R. Millán, "Feature extraction for multi-class BCI using canonical variates analysis," in *Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on*, 2007, pp. 1–6.

[37] E. M. Holz, J. Höhne, P. Staiger-Sälzer, M. Tangermann, and A. Kübler, "Brain–computer interface controlled gaming: Evaluation of usability by severely motor restricted end-users," *Artificial Intelligence in Medicine*, vol. 59, no. 2, pp. 111–120, 2013.

[38] A. Kübler, F. Nijboer, J. Mellinger, T. M. Vaughan, H. Pawelzik, G. Schalk, D. J. McFarland, N. Birbaumer, and J. R. Wolpaw, "Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface," *Neurology*, vol. 64, no. 10, pp. 1775–1777, 2005.

[39] D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw, "Brain–computer interface (BCI) operation: Optimizing information transfer rates," *Biological Psychology*, vol. 63, no. 3, pp. 237–251, 2003.

[40] D. Berrar, I. Bradbury, and W. Dubitzky, "Avoiding model selection bias in small-sample genomic datasets," *Bioinformatics*, vol. 22, no. 10, pp. 1245–1250, 2006.

[41] G. Müller-Putz, R. Scherer, C. Brunner, R. Leeb, and G. Pfurtscheller, "Better than random: A closer look on BCI results." *International Journal of Bioelectromagnetism*, vol. 10, no. 1, pp. 52–55, 2008.