IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 26, 2020, accepted January 8, 2021, date of publication March 8, 2021, date of current version March 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3064226

Post-Adaptation Effects in a Motor Imagery Brain-Computer Interface Online Coadaptive Paradigm

JOSÉ DIOGO CUNHA[1], SERAFEIM PERDIKIS[2], (Member, IEEE), SEBASTIAN HALDER[2], AND REINHOLD SCHERER[1,2], (Member, IEEE)
[1]Laboratory of Brain-Computer Interfaces, Institute of Neural Engineering, Graz University of Technology, 8010 Graz, Austria
[2]Brain-Computer Interfaces and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, U.K.

Corresponding author: Reinhold Scherer (r.scherer@essex.ac.uk)

ABSTRACT Online coadaptive training has been successfully employed to enable people to control motor imagery (MI)-based brain-computer interfaces (BCIs), allowing to completely skip the lengthy and demotivating open-loop calibration stage traditionally applied before closed-loop control. However, practical reasons may often dictate to eventually switch off decoder adaptation and proceed with BCI control under a fixed BCI model, a situation that remains rather unexplored. This work studies the existence and magnitude of potential post-adaptation effects on system performance, subject learning and brain signal modulation stability in a state-of-the-art, coadaptive training regime inspired by a game-like design. The results extracted in a cohort of 20 able-bodied individuals reveal that ceasing classifier adaptation after three runs (approx. 30 min) of a single-session training protocol had no significant impact on any of the examined BCI control and learning aspects in the remaining two runs (about 20 min) with a fixed classifier. Fifteen individuals achieved accuracies that are better than chance level and allowed them to successfully execute the given task. These findings alleviate a major concern regarding the applicability of coadaptive MI BCI training, thus helping to further establish this training approach and allow full exploitation of its benefits.

INDEX TERMS Brain-computer interface, classifier adaptation, coadaptation, motor imagery, online learning, user training.

I. INTRODUCTION

Brain-computer interfaces (BCI) based on the detection and identification of electroencephalographic (EEG) sensorimotor rhythms (SMRs), which are elicited by imagined or attempted movements [1], are popular for providing the possibility of spontaneous interaction by noninvasive means. Entering a phase of considerable technical maturity, the applicability of the motor imagery (MI) BCI paradigm has been demonstrated in several contexts including communication [2]–[4], games [5], [6], virtual worlds [7], prosthetic devices [8], [9], brain-actuated wheelchairs [10], [11], mobile robots [12], robotic arms [13] and rehabilitation [14], [15].

Intense performance fluctuations during and, especially, across BCI sessions [16] and the inability of a large portion of users to get into control of a BCI have been early identified as [17], and still remain today [18], the main obstacles towards deploying BCI technology in real-world scenarios [19]–[23]. SMR-based BCIs are known to be particularly vulnerable to these issues [18], [24]–[27]. Coadaptive BCIs, where the decoder parameters [28] and/or–less commonly–the features extracted from brain signals to be processed by the decoder [29]–[31] are recalculated on-the-fly during real-time, closed-loop BCI operation have been proven able to reduce performance instability by tracking and adapting to non-stationarity effects present in brain signals [16], [27]–[30], [32]–[38].

This literature has often implied that decoder adaptation subserving online BCI control may also pose a remedy for the problem of non-universal accessibility (often termed "BCI illiteracy" [34]), when subjects do not exhibit the desired modulation of SMRs to be exploited by the BCI. Although

The associate editor coordinating the review of this manuscript and approving it for publication was Ludovico Minati.

it is still unclear and debatable whether and to what extent an evolving decoder enhances or hinders subject learning on the long run [6], [39], [40], fast or even immediate transition to closed-loop training is thought to be beneficial. In that respect, adaptive BCIs do provide a means to avoid open-loop (i.e., without BCI feedback) calibration protocols altogether. This has been so far a necessary training stage in all modern BCI paradigms relying on machine learning, since the employed algorithms prerequisite a supervised parameter estimation process demanding the synchronized collection of both brain signal data and ''ground truth'' mental task labels. In other words, there is need of knowledge about the underlying mental tasks which is not always readily available. However, ''offline'' data collection is a lengthy, tedious and dull procedure that largely demotivates prospective BCI users. Given compelling evidence on the importance of motivation [41], [42] and the highlighted need to take into account psychological factors in training protocol design [43], [44], coadaptive schemes are believed to be able to play a major role in BCI training.

There exist several possible reasons why BCI adaptation may need to cease and subsequent training or BCI application control proceed with fixed decoders. One such case may be when trained performances with a supervised adaptation paradigm are judged adequate for application control, but the targeted BCI device does not provide data labels to permit continuing adaptation in the same supervised manner. Of note, in spite of successful adoption in specific frameworks and under certain valid assumptions [27], [29], [30], [34], unsupervised re-calibration schemes are not guaranteed to converge and can be in cases theoretically shown to fall short of their supervised counterparts [38]. Computational complexity during brain-actuated device operation may also pose limitations on continuous adaptation. Most importantly, online parameter estimation is often employed only in the beginning of BCI sessions to address non-stationarity effects and then deliberately switched off [39], [45] since it is widely believed that fixed decoders may more effectively foster subject learning and fuel the associated cortical plasticity [6], [8], [39], [40], [46]–[49].

Despite post-adaptation operation may be a practical necessity and although this issue has been investigated with respect to other BCI paradigms [39], [45] the literature of adaptive SMR BCIs has so far studied performances only during, and not after, adaptation. Faller *et al.* [37] is, to our best knowledge, the only SMR BCI work that has included closed-loop control intervals devoid of adaptivity following coadaptive training; however, this experimental design served to investigate a different hypothesis and an explicit comparison of performances in the two conditions was not attempted. In [50], [51], the authors simulate offline the performances of adaptive and static classification schemes arguing in favour of the former, however, no firm conclusions can be drawn in the absence of actual closed-loop experimentation.

In this work we set out to explicitly study the presence and magnitude of effects in terms of system performance, subject learning and feature stability, following the discontinuation of adaptation in a coadaptive training regime. We maintain an exploratory attitude in this investigation, where no specific hypotheses on the nature of potential effects are put forward. This is due to the fact that there exist several lines of reasoning suggesting that adaptation and its stoppage may influence variables of interest (including the stability of brain features, subject learning and, ultimately, the overall system performance), but these are often contradictory. For instance, as far as stability is concerned, most works imply that non-stationarity is inevitable and adaptation helps reduce its negative impact on performance. However, it is also sensible to suspect that a continuously changing (due to adaptation) feedback may in fact enhance non-stationarity effects [6], [38]. Similarly, as already mentioned, there is an open debate about the impact of adaptivity on subject learning [40]: On the one hand, higher performances thanks to adaptation translate in more meaningful and less frustrating feedback, which can be assumed to facilitate learning. On the other hand, continuous adaptation leads to inconsistent feedback creating the so-called ''moving target problem'' [46], [52] where users need to learn an ever-changing task, which should be detrimental to their learning efforts [6], [39]. It has also been argued that continuous adaptation may be counterproductive for subject learning by means of over-facilitating the underlying task, so that ''lazy'' human learners may not achieve the desired outcomes [38]. Consequently, the final effects on BCI performance will depend on which of the above positions prove to be prevalent; this calls for online, closed-loop experimentation.

In order to explore these effects, a cohort of 20 BCI naive, able-bodied individuals have been recruited and undergone a single-session, 5-run-long coadaptive training protocol. Each run consisted of 40 hand MI trials. Except for the first 10 trials of the first run that were used to calibrate the initial instance of the BCI model, users were in closed-loop control of the interface and observing real-time feedback throughout the training session. The users' task was to play a version of the ''Whack-A-Mole'' game [53] where timely, successful and sustained MI would ''hammer'' and knock out a cartoon ghost character leading to a ''hit'' trial and collection of ''stars'' (virtual rewards). The game-like design targeted high levels of user motivation and engagement to facilitate coadaptation [54], [55]. Feedback relied on SMR feature classification. In line with the research question addressed, after initial calibration the classifier was recurrently adapted in a supervised manner during the first 3 runs. In the subsequent and latest two runs of the session, adaptation was switched off and users continued closed-loop training with the last classifier resulting from the adaptive phase. At the end of each training stage, subjects were asked to subjectively assess their feeling of control, satisfaction with the training paradigm, level of alertness and performance evolution. Variables related to user effort and comfort were reported once at the end of the session with a simplified variant of the NASA-TLX workload assessment tool (see supplementary material).

The remainder of this manuscript is organized as follows: Section II elaborates the details of the participants recruited, the experimental apparatus, the training paradigm applied, the BCI, evaluation and statistical methods employed, and the data usage. Section III presents the system performance, subject learning, stability and user experience results substantiating our claim that there exist no significant post-adaptation effects in SMR BCI training paradigms. Finally, section IV discusses our findings with respect to related literature and highlights their significance and anticipated impact.

II. METHODS

A. PARTICIPANTS

Twenty naïve volunteers (mean age 26 ± 3 (s.d.) years, 5 females, 2 left-handed–S8 and S15) participated in this study. Subjects were without any known medical condition, had normal or corrected to normal vision and entered the study voluntarily without monetary remuneration. At the beginning of the study, each participant was briefed about its aim. All volunteers gave written informed consent to participate. The study was conducted in accordance with the relevant guidelines for ethical research set by the Declaration of Helsinki.

B. EXPERIMENTAL APPARATUS

Recording sessions took place in a well illuminated, spacious room shared by the subject and 2 researchers. Subjects were seated in a comfortable chair approximately 90 centimeters from the computer monitor. EEG signal was recorded at 512 Hz sampling rate with an eegosports biosignal amplifier (ANT Neuro, Enschede, Netherlands) from 15 Ag/AgCl electrodes over locations FC3, FCC1h, FCC2h, FC4, C5, C3, C1, Cz, C2, C4, C6, CCP1h, CCP2h, CP3, and CP4 of the sensorimotor cortex according to the international 10-20 system (Fig. 1a). These electrodes were integrated into a customized 64-channel cap (waveguard, ANT Neuro, Enschede, Netherlands). All signals were referenced to electrode CPz and kept with impedance under 20kΩ. The ground electrode was placed at AFz.

C. TRAINING PARADIGM

The BCI system was presented to the user in the form of a game, so as to maintain motivation [43], [44]. Participants were seated in front of a computer screen with the task to play a variant of the Whack-A-Mole game. Users were asked to perform dominant hand MI whenever a cartoon "ghost" character was shown on the screen and to relax when it disappears. The ghost would appear, play an audio cue and stay visible for 7 seconds, which constitutes a single MI trial. A break between 7 to 8 seconds was presented before the next ghost appeared, which defines a single non-MI (rest/idle) trial. The trial timeline is shown in Fig. 1b.

Prior to training, participants were given instructions on the ideal execution of kinaesthetic hand MI tasks [56]. Each subject was given a lemon to squeeze and to feel its texture. Then, they were asked to mentally reproduce the motor and haptic feeling of squeezing the lemon (not the visual image of doing it) with their dominant hand. They were advised to pick a strategy that feels vivid when imagined, and that they should keep it consistent for the next hour, since the classifier works at its best with consistent strategies. After this, subjects were informed about EEG artifacts and how they affect the data. By means of an EEG signal scope, subjects were able to see for themselves how muscular contractions and eye blinking affect the signal. Finally, participants were asked to avoid blinking with a pattern (e.g. blinking every time the ghost appears), but to otherwise blink naturally, as needed.

The graphical user interface (GUI) of the game and the trial timeline are illustrated in Fig. 1b. The game GUI is composed of several elements designed to provide rich feedback to the user while keeping the training paradigm appealing and straightforward. Specifically, during resting trials, a hammer is depicted in a vertical, upright position ("Resting" position). While the BCI output indicates that the user is idling (i.e., not performing MI), a blue bar on the left of the hammer keeps filling up (and becoming a lighter blue color) showing the "energy" recovered by the user. The blue bar reaches maximum energy after 5 s of continuous resting. Every 20% of energy recovered earns the subject one star, as indicated by the next "empty" (white) star (in the 5-start-high column next to the energy bar) becoming blue. This type of feedback promotes the user's intentional-non-control abilities [19], which are essential for self-paced BCI (i.e., spontaneous control interfaces where the users initiate BCI commands at their own will and pace), as well as helps to improve the main control abilities, since a consistent "resting" EEG pattern can be optimally separated from a user's MI pattern. Effectively, the user is learning to avoid false positives, since intervals during a resting trial where the BCI detects false-positive MI activity prevent the energy recovery process and negatively affect star collection. During MI trials, when the BCI infers that the user is engaging in MI for more than 0.3 seconds continuously, the hammer increases in size and prepares to hit the ghost ("Charging state"). No energy is lost or recovered in this state. When continuous MI detection exceeds 2 s, the hammer strikes and hits the ghost ("Hit"), upon which the user capitalizes the currently available stars and the energy resets to 20% (1 star remains blue). At the end of MI trials (ghost disappearance), the energy would reset to 0% (all stars are empty/white). The time counter for charge and hit resets every time the BCI fails to detect MI within the MI trial. Given the total duration of 7 s for MI trials and a minimum of 2 s for a single hit, subjects have the possibility of a maximum of 3 hits in each trial, although each subsequent hit after the first one can only contribute one star to the total reward count. The gameplay within MI trials is meant to encourage the fast and persistent production of MI EEG activity. The total number of stars collected by the user in a training run is displayed at the bottom of the screen.

Each recording session lasted about 90 min (including EEG montage, instructions, recording and periodic pauses)

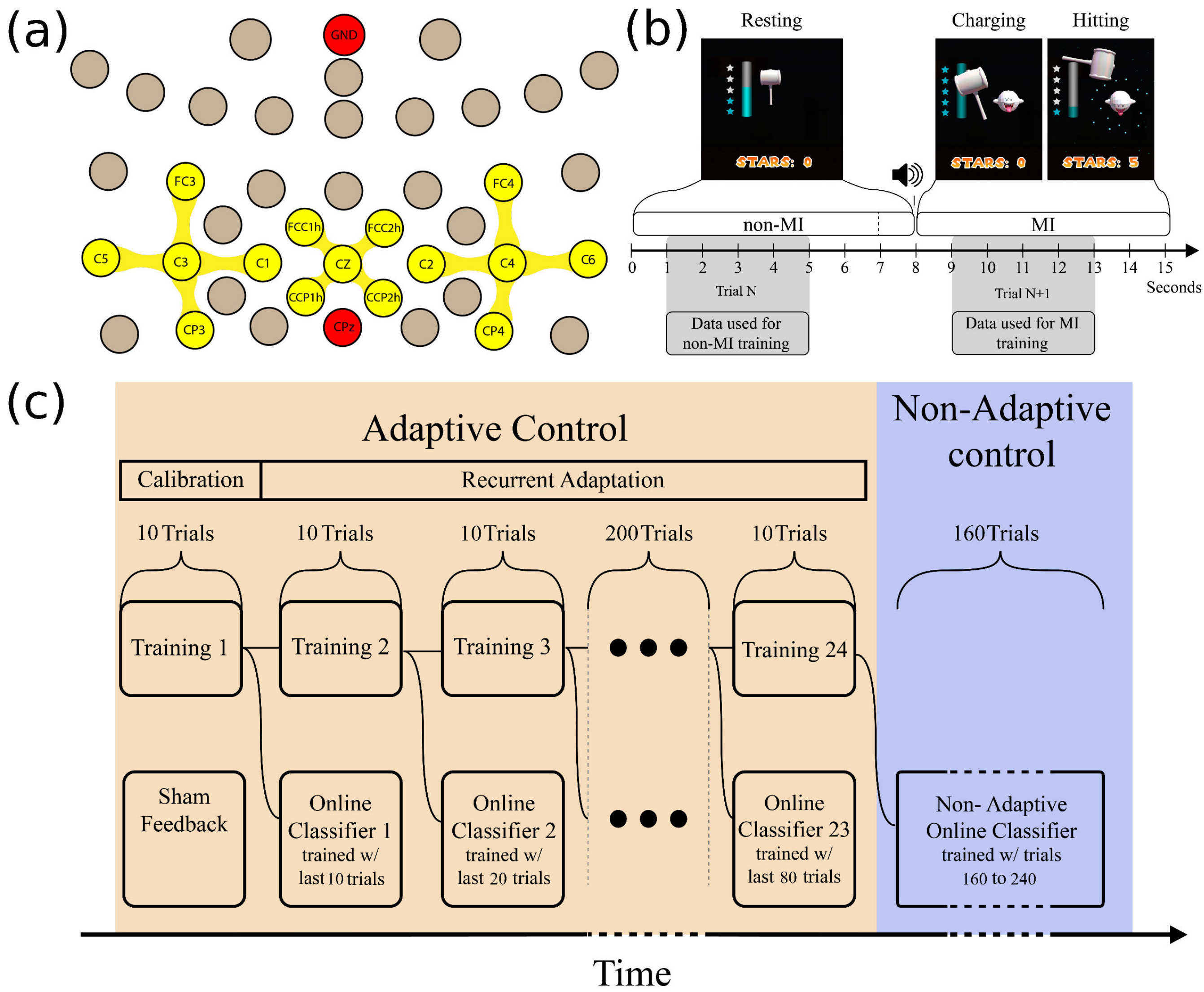

FIGURE 1. **Methods for MI BCI coadaptive training. (a) EEG channel configuration over the sensorimotor cortex according to the international 10-20 system. Electrode CPz acted as reference and AFz as ground. (b) Game graphical user interface associated to the trial structure and timeline. Image adopted from [53]. (c) Training paradigm structure. Each MI trial is preceded by a non-MI, "resting" trial as shown in (b).**

and consisted of 5 runs of 80 total trials each (40 MI trials, each preceded by a rest trial). Between runs, subjects were free to take a break to move, ask questions, drink and eat. The duration of this pause was decided by each subject and would take approximately 1-5 min.

In the first 3 runs (240 total trials) BCI model adaptation was enabled. The online coadaptive training was divided in calibration and recurrent adaptation [36]. The aim of the calibration period is to collect a minimal amount of only 10 EEG trials to compute the first set of BCI parameters. During calibration, sham feedback is provided to the user, i.e. the game is playing automatically with a predefined accuracy of 75% (open-loop BCI with sham feedback). As soon as this first version of a calibrated BCI is available, training switches to the main recurrent adaptation stage, where the user is in direct control of the interface (closed-loop BCI). In this recurrent adaptation phase, a new classifier is seamlessly trained every 10 total trials (5 MI + 5 non-MI trials) in the background, without disrupting the flow of the training protocol. Each such classifier is trained on the data of the last 80 trials (or, if less than 80, all currently available trials), in order to give recent activity patterns a higher impact. Only data in time intervals [1,4] s within each MI trial and [−5, −2] s within each preceding rest trial (where $t = 0$ the MI trial onset, when the ghost appears) are used for the classifier update, so as to filter out potentially bad quality data either in the beginning (delayed reaction) or at the end (inability to sustain MI) of an MI trial. After the end of the recurrent adaptation stage, subjects executed another 2 runs (160 trials) without adaptation, using the last classifier resulting from the adaptive stage. The net recording time was about 50 min (30 min for the adaptation stage and 20 min for the fixed classifier stage). The training paradigm is graphically displayed in Fig. 1c.

All subjects were asked to report their feeling of control, satisfaction, alertness and performance improvement twice, after the end of each training stage, as well as to declare various overall user experience variables at the end of the session with a simplified version of the NASA-TLX workload assessment tool (see supplementary material for the exact questionnaire form used). Self-reported level of control was assessed with the question "Did you feel in control of the game in this stage?", which could be answered with an integer number in the scale 1-10 (1: Very little - 10: Very much). User satisfaction was evaluated through the question "What is your overall satisfaction with the system?", where participants could reply again in the scale 1-10 (1:Unsatisfied - 10:Satisfied).

D. BCI METHODS

For online MI BCI, a large Laplacian derivation [6], [7], [38] was first applied to channels C3, Cz and C4, where the average scalp potential of their 4 "cross" neighbor channels (North-South-East-West) was subtracted from the raw EEG signal at each time point. Subsequently, the signal of these 3 Laplacian channels was filtered throughout the 3 s-long interval of each trial with 3 non-causal, 5^{th}-order, bandpass Butterworth filters setting the cutoff frequencies at 10-13, 16-24 and 24-32 Hz (α, low β and high β / low γ bands, respectively). The final features were extracted in 1 s-long consecutive windows shifted by 125 ms. A bandpower estimate for each of the three filtered versions of each window is derived by squaring the signal, averaging across time and log-transforming the output (in order to increase normality of brain features). As a result, the final feature set the participants were asked to modulate consisted of 9 subject-unspecific features: 3 log-power values corresponding to the μ, low β and high β bands on the aforementioned 3 Laplacian channels.

These 9-dimensional feature vectors are classified by a shrinkage-regularized [57] linear discriminant analysis (LDA) classifier. The parameters of each version of the LDA classifiers applied online are extracted by conventional maximum-likelihood estimation of class-dependent mean vectors and a common (regularized) covariance matrix using the recent available data as described in the previous section. Such parameter estimation is of low computational complexity and was instantaneously and seamlessly implemented at the specified intervals without disrupting the flow of the training protocol, thus being a process entirely concealed from the user. Prior to each classifier recalibration step, an artifact rejection module was employed to detect and remove from the training dataset feature vectors that were suspected for artifact contamination. Specifically, a normal distribution was first fitted to the recalibration data of each of the 9 features used, independently. Feature values at any time t whose absolute z-score exceeded the threshold 3.0 (i.e., feature values whose log-power was 3 standard deviations larger or smaller than the average feature value in the training data) were labeled as abnormal and the overall feature vector/data sample at time t was rejected and ignored by the classifier retraining process. Post-experiment analysis revealed that artifact infliction incidents in the overall study were in fact rare, so that no artifact rejection module was implemented for the offline analyses.

The classification was driving the hammer feedback as follows: Instead of the common "hard" classification approach based on the classifier's separation hyperplane, the current LDA classifier estimate was used to derive for each incoming feature vector a "soft" decision in the form of a probability distribution over the two mental classes (rest, MI). When the probability of the MI class exceeded 0.55, the sample-wise classifier inference on the user's current mental state was taken to be "MI engagement". Sustaining this state for the preset intervals mentioned in the previous section would lead to the hammer's charging and hitting actions. On the contrary, when the MI probability was below the 0.55 threshold the user's state was interpreted as "Resting". A resting single-sample classification decision would reset the counters for the hammer's charging and hitting states, so that strong and sustained MI was key to meeting the game's goals.

The selection of this particular threshold (instead of the more conventional 0.5) was meant to slightly bias decision making towards the avoidance of false positives and challenge participants to put more effort. Perfect control would account for a total of 280 stars gained per run. A random classifier was estimated to deliver approximately 64 stars per run. Classification fully biased towards the resting class would lead to 0 stars gained, whereas bias to the MI class would only deliver 80 stars per run (thanks to the "energy recovery" mechanisms during resting trials). Of note, the last two scenarios of biased classification were practically completely avoided thanks to recurrent adaptation.

Following the 125 ms window shift for feature extraction and classification, the rate of single-sample decision making was 8 Hz, however, the subsequent evidence accumulation approach through which single-sample decisions were driving the hammer feedback allowed for more relaxed gameplay dynamics which were assumed to better promote learning.

E. EVALUATION METHODS

We report the online single-sample classification accuracy metric in the conventional fashion (for the purpose of comparison with previous work) by pulling together and thresholding the MI probabilities produced online by each subject at 8 Hz within the trial periods with a 0.5 threshold (including samples whose rising edge of 1-s-long window precedes the trial onset, since the resulting feature vectors were taken into account during real-time operation). These "hard" (MI vs rest) single-sample decisions are used to form standard confusion matrices (row-wise, True Positives (TP): Number of correctly classified MI samples, False Positives (FP): Number of rest samples misclassified as MI, False Negatives (FN): Number of MI samples misclassified as rest, True Negatives

(TN): Number of correctly classified rest samples). The final accuracy metric ($Accuracy = 100 * (TP + TN)/(TP + FP + FN + TN)\%$) reflects the percentage of correctly classified samples over the total number of samples (including both rest and MI trials).

An online trial-based (trial-wise) accuracy metric is derived in the same fashion considering a single final decision per MI and rest trial. This decision is calculated by simple majority voting of single-sample decisions within each trial (using again the 0.5 threshold for the latter). Using the same trial-wise confusion matrix, the Hit percentage ($Hits = 100 * TP/(TP + TN)\%$) is presented to isolate performance during MI trials only. Similarly to the single-sample accuracy metric, these evaluation metrics do not precisely reflect the online performance perceived by the subjects during training through the employed gameplay, however, they are closer to standard practice so as to allow comparisons with the literature. Finally, the average number of "star" rewards actually collected by each participant per run is reported to give a sense of the performance actually observed by subjects. Of note, all accuracy indices are monitored per run, condition or session, by pulling the corresponding data together (instead of averaging run-wise accuracy within the respective periods), unless otherwise stated.

Event-related Desynchronization/Synchronization (ERD/ERS) maps of MI trials are derived with the classical definition [1] expressing a decrease/increase of bandpower in a certain location, time point and frequency band over a reference interval. The reference period is taken to be the interval $[-3, -1]$ s within the preceding rest trial with respect to the onset of each MI trial ($t = 0$). Each trial is processed with Laplacian spatial filtering using all existing surrounding neighbours and DC removal prior to ERD/ERS calculation. The EEG spectra are computed with Fast Fourier Transform (FFT) with a frequency resolution of 0.125 Hz in the broad band $[8, 36]$ Hz. Time resolution is 125 ms. The final map shown for each channel is the average spectro-temporal ERD/ERS across all MI trials of all 5 runs. For topoplots illustrating the spatial distribution of ERD/ERS, the signal is shown for all available channels, further averaging across the whole MI trial period and within the noted frequency sub-band.

Several indices are extracted to assess the separability between the subjects' MI and resting EEG patterns and, by extension, their learning outcome. Offline single-sample accuracy is computed in the same manner as the online equivalent except for the fact that single samples are classified by non-regularized LDA classifiers trained with 3-fold cross validation. Random shuffling of data is applied, but samples originating in the same trial remain in the same fold to avoid overestimating accuracy due to data sample dependence. The final reported figures refer to average across folds testing set accuracy. The coefficient of determination r^2 [29] is computed as the square of the Pearson correlation coefficient (assuming gaussianity of brain features) between feature values and the corresponding ground truth mental class labels, separately for each of the 9 features employed online. The final figures shown are those corresponding to the "best" (in terms of r^2) feature of each subject considering the whole session. Similarly, Fisher Score (FS) [6] values are also reported for the best feature session-wise using the formula in (1), where μ, s the mean and standard deviation of the feature in question for the MI and resting classes. Fisher Score is an intuitive way to assess the extent of overlapping (thus, the proximity/similarity) between univariate normal distributions and it is theoretically associated to the t-statistic.

$$FS = \frac{|\mu_{MI} - \mu_{rest}|}{\sqrt{s_{MI}^2 + s_{rest}^2}} \quad (1)$$

Finally, Kullback-Leibler divergence (KLD) separability [38] is computed as in (2), where f, Σ, $\boldsymbol{\mu}$ the (assumed normal) multivariate distribution probability density function, the covariance matrix and the mean feature vector of the two mental classes (MI and rest) and $D = 9$ is the dimensionality of the feature space. KLD is commonly adopted as a measure of distance and similarity between probability distributions.

$$\begin{aligned} KLD &= D_{KL}(f_{MI}||f_{rest}) \\ &= \frac{1}{2}(tr(\Sigma_{rest}^{-1}\Sigma_{MI}) \\ &\quad +(\boldsymbol{\mu_{MI}} - \boldsymbol{\mu_{rest}})^T \Sigma_{MI}^{-1}(\boldsymbol{\mu_{MI}} - \boldsymbol{\mu_{rest}}) \\ &\quad -D - \ln\frac{|\Sigma_{rest}|}{|\Sigma_{MI}|}) \end{aligned} \quad (2)$$

Of note, all separability indices employed here depend to a certain extent on the assumption of normal univariate (for individual features as for r^2 and FS) and multivariate (for the 9-dimensional space used for offline accuracy and KLD) distributions. Thanks to log-transforming the extracted bandpower features the data, this assumption is sufficiently accommodated in the case of this work for all 9 features used online across all subjects. As already noted, all these metrics are computed on the basis of the same feature manifold used for online BCI.

It must be underlined that these separability metrics are complementary in nature and each has its own pros and cons, what motivates the inclusion of all these aspects of class discriminancy in our analysis. Offline accuracy is probably the most easily interpretable and intuitive metric, it can evaluate the separability of multivariate patterns (without having to average or take the maximum over features), it is numerically bounded on both sides and enjoys clear theoretical estimates about the expected performance in non-separable datasets. On the downside, it is an indirect measure of separability that relies on the classification model chosen, a large number of hyperparameters (e.g. model-specific parameters, number of fold for cross-validation, etc) and is prone to computation errors like overfitting, underfitting and assumption violations (normality, independence). Furthermore, classification accuracy values tend to saturate close to the chance level for little-separable datasets and exhibit large local variability, thus not allowing to differentiate the aptitude of low-performing users, which is usually the most interesting subject group in learning studies. r^2 is also bounded on both sides, but it can only be computed per feature and it is not very intuitive since the absolute values tend to be specific to the feature space employed. Offline accuracy and r^2 may further be considered rather ''lenient'' measures of subject learning, since improvements may reflect slightly increased percentages of samples lying on the ''correct'' side of the separation hyperplane, which however probably do not represent considerable deviations of the feature distributions and thus, actual skill learning. Fisher Score is a direct measure of distribution separability, however, it is also not easily interpretable, it can only be computed for individual features and it is unbounded from above. Finally, KLD is the only other metric next to accuracy that can assess multivariate datasets with a single attribute, but, it is unbounded from above and particularly sensitive to the number of available samples (because of the need to estimate full, class-wise covariance matrices).

The stability of the MI distribution (equivalently, the magnitude of potential non-stationarity effects) is computed by means of the same KLD definition, where the probability density function distance is taken between the MI distribution estimates in two consecutive runs (rather than between the MI and rest class distributions in the same chunk of data, as for the KLD separability metric). Feature stability assessment is based on the extraction of a new set of bandpower features on higher spatial and spectral resolution so as to allow eventual instabilities to emerge. Specifically, after spatially filtering EEG channels with a local Laplacian derivation taking into account all surrounding neighbours and removing their DC component, we calculate the power spectral density (PSD) in each of the 15 Laplacian channels between 8-30 Hz with 2 Hz resolution over the last second. The PSD is computed every 62.5 ms (i.e., 16 times per second) using the Welch method (five 25%-overlapping internal Hanning windows of 500 ms) and log-transformed. The final feature vectors are thus spectral density estimates on combinations of 15 channels and 12 frequency bands for a total of 180 features. The features within this new manifold are ranked according to discriminant power by means of Fisher Score (equivalent results are obtained when using r^2) for each individual run. Finally, the feature stability index is derived by computing the cardinality of the intersection between the best-10 feature sets in two consecutive runs, S_k, S_{k+1} (3).

$$FeatureStability = \frac{|S_k \cap S_{k+1}|}{max\{|S_k|, |S_{k+1}|\}} = \frac{|S_k \cap S_{k+1}|}{10} \quad (3)$$

F. STATISTICAL TESTING

Average or median values across the specified groups are selected as point estimates for the various metrics reported in our analysis. Dispersion is shown either through standard deviations (s.d.) or ranges between the 25th and 75th percentile (boxplots). Linear correlations are reported by means of the Pearson correlation coefficient and its significance at the 95% confidence interval through the corresponding Student's t distribution. Statistical differences between any two populations (i.e., comparisons between adaptation conditions or experimental runs) are assessed with non-parametric, paired, two-sided Wilcoxon signed-rank tests (so as to avoid issues with potential non-gaussianity of the underlying data) at the 95% confidence interval. In marginal cases, the p-values of two-sided, paired t-tests are also provided. No correction for multiple comparisons is attempted in the light of the fact that almost all effects monitored are not statistically significant without correction. Unless otherwise stated, in order to assess the existence of increasing or decreasing trends in the inspected metrics between two time points t and t', we opted for paired tests between the respective populations, instead of the common alternative involving testing the hypothesis that the distribution of individual paired differences differs from the zero-mean distribution. However, the two processes are largely equivalent and the conclusions reached here are not affected by this choice.

G. DATA USAGE

No participant has been removed from the analysis reported in this manuscript. All results are based either on saved raw EEG data or processed feature vectors and class-posterior probabilities extracted during online operation and saved on the fly. The raw EEG data of subject S10 were lost due to technical problems. Consequently, this subject is excluded from any result relying on these data, namely, the ERD/ERS topographic maps and the feature stability outcomes. Four subjects (S8, S10, S12, S14) executed an additional one or two sessions with the same protocol. This additional data were not taken into account for any analysis presented here.

III. RESULTS

A. COADAPTIVE TRAINING

Showcasing that the implemented coadaptive training paradigm can successfully bring users in control of the BCI verifying the state-of-the-art is a necessary foundation for investigating post-adaptation effects. Hence, we first present BCI performances and confirm that these are accompanied and driven by the anticipated spatiospectral SMR patterns.

1) BCI PERFORMANCES

Table 1 illustrates the single-sample classification accuracy, the trial-wise accuracy (considering a single yes/no MI-detection decision per trial–including resting trials–which is extracted through majority voting of single-sample decisions within the trial), the trial hit percentage (true positives of trial-wise confusion matrix normalized by the total number of MI trials in the session) and the total number of collected stars. All metrics are computed session-wise pulling all data together. Performances are sorted in descending order by single-sample accuracy.

Given the adequate number of trials recorded ($N = 200$ per class) and single-samples (at least $N = 11200$ per class)

TABLE 1. Online performance throughout the training session. Accuracy: single-sample classification accuracy. Trial Accuracy: trial-wise accuracy. Hits: True Positive percentage of trial-wise accuracy. Stars per run: collected stars averaged over runs.

Subject ID	Accuracy (%)	Trial Accuracy (%)	Hits (%)	Stars per Run
S1	78.3	95.4	98.5	222
S2	72.9	89.5	93.4	203
S3	68.1	83.7	84.7	182
S4	67.6	79.3	87.8	197
S5	67.5	85.2	83.7	179
S6	65.5	77.8	96.9	171
S7	63.9	73.5	71.9	161
S8	62.7	76.3	87.8	163
S9	60.3	64.5	93.9	156
S10	59.7	66.8	70.9	109
S11	58.5	65.6	54.1	99
S12	57.9	65.3	62.8	108
S13	57.8	63.3	66.8	124
S14	56.4	61.0	59.2	124
S15	55.9	58.4	55.6	67
S16	54.6	56.1	69.9	92
S17	53.3	53.8	41.8	52
S18	52.8	51.5	80.6	94
S19	51.8	49.5	42.9	50
S20	50.7	50.5	63.8	63
Mean ± s.d.	60.8±7.4	68.4±13.6	73.3±17.5	132.4±48.3

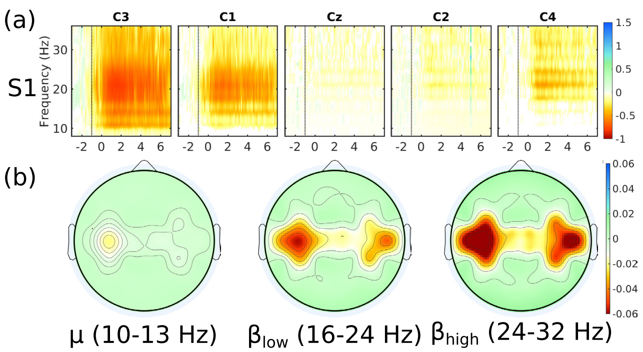

FIGURE 2. Evidence of sensorimotor rhythm modulation during coadaptive training. (a) ERD/ERS maps of best performer S1 (right-handed) throughout the session. (b) Average (across subjects) topographical distribution of ERD/ERS within the μ (10-13 Hz), low β (16-24 Hz) and high β (24-32 Hz) bands.

in a session, the chance level approaches 58% at the 95% confidence interval assuming binomial distribution of decisions for both single-sample and trial accuracy [58]. Therefore, 13 (sample-wise) to 15 (trial-wise) out of 20 subjects can be said to have acquired at least a minimum necessary level of control over the interface thanks to coadaptation, despite directly starting with online control in a single session. Nine subjects (S1-S9) achieved performances well above chance level. This outcome is consistent with the corresponding literature [27]–[30], [34], [36]–[38]. The percentage of hit trials is supportive of the same conclusion. Furthermore, strong, significant correlations between collected stars and (both) single-sample and trial accuracy ($r = 0.94$, $p < 10^{-9}$, $N = 20$, in both cases) suggest that the game-like motivational paradigm adopted has been successful in precisely rewarding performance while maintaining motivation.

Fig. 2 shows that the aforementioned BCI performances come as a result of the anticipated cortical activation elicited by the subjects' MI, as previously established in a large body of literature that has studied the neurophysiology and the EEG correlates of imagined motor tasks [1], [16], [36]. Specifically, Fig. 2a showcases the exemplary event-related desynchronization/synchronization (ERD/ERS) [1] maps of the best performing subject (S1) over 5 central, bilateral and medial locations of the sensorimotor cortex. As expected, MI manifests mainly as an ERD (drop in bandpower with respect to the reference interval before trial onset $t = [-3, -1]$, red color) approximately time-locked to the onset of MI ($t = 0$) and largely sustained until the end of trials. Additionally, ERDs are particularly dominant in the μ and β bands as previously described. Finally, given that all subjects were instructed to use their dominant hand and that only two participants were left-handed (S8, S15), despite the presence of ipsilateral SMR modulation (channel C4), the evident predominance of contralateral MI correlates (channels C1,C3) is also consistent with the literature. Fig. 2b illustrates the scalp distribution of ERD/ERS in the three bands of interest, averaged across trial duration and the overall subject population, confirming that sound SMR cortical patterns (i.e., bilateral ERDs in μ and β with a predominantly contralateral component with respect to right hand imagery) were observed throughout the recruited population. Importantly, it is not claimed that these average ERD/ERS topographic maps represent a generic brain pattern of the MI pair used by the participants, but, rather, that the aforementioned population trends extracted in the literature are verified. The fact that all subjects exhibited physiologically relevant brain patterns, in spite of the expected subject specificity, is detailed in supplementary Fig. S1 depicting the topographical distributions of ERD/ERS individually for each participant.

The benefits of such a coadaptive training regime cannot be limited to its effectiveness in bringing the expected percentage of prospective users in control of the BCI, since it is established that this milestone can be also achieved by conventional training protocols, i.e., those imposing a longer open-loop calibration procedure before closed-loop control. Hence, the real added value of coadaptation is that closed-loop BCI is enabled, essentially, immediately (i.e. after only 10 open-loop trials), thus directly engaging the learning abilities of both the subject and the machine, while maintaining the interest of human participants in the training procedure. In that respect, it is shown (by simulation) that a conventional calibration approach using the first 3 runs for BCI model training and applied to the remaining two runs would result in average single-sample classification accuracy of $61.4 \pm 7.6\%$, as opposed to $61.2 \pm 7.7\%$ derived online by the proposed coadaptive training scheme. Hence, skipping the undesired open-loop calibration comes at no detriment whatsoever to the BCI parameter estimation procedure. The fact that coadaptation does not seem to improve, in terms of accuracy, over open-loop training either, is elaborately discussed in Section IV. Furthermore, as shown in Fig. 3a, the BCI model coming out of the 10-trial calibration with sham feedback significantly underperforms compared to the online accuracy for both the adaptive ($p < 10^{-4}, N = 20$) and static (non-adaptive) runs ($p < 10^{-3}, N = 20$, two-sided, paired, Wilcoxon signed-rank tests). In other words, a very short open-loop recalibration alone is not sufficient towards a positive training outcome; the subsequent coadaptation process is necessary for BCI control.

2) USER EXPERIENCE

Self-reported level of control over the BCI and user satisfaction with the training paradigm further support its effectiveness (Table 2). The vast majority of subjects (17/20) assessed their control capacity with a grade above 5 in a scale from 1 to 10 in at least one of the two stages, a result consistent with the number of subjects performing better than a random classifier, whereas the average user satisfaction taking into account both stages was 7.4 in the same scale. Notwithstanding the inevitable subjectivity of questionnaire-based evaluation, strong and statistically significant correlations between single-sample accuracy and control in the respective stages ($r = 0.78$, $p < 10^{-4}, N = 20$ during recurrent adaptation and $r = 0.58$, $p = 0.008, N = 20$ for the fixed BCI stage), as well as with satisfaction ($r = 0.52$, $p = 0.02, N = 20$ during recurrent adaptation and $r = 0.49$, $p = 0.0029, N = 20$ for the fixed BCI stage) substantiate the relevance of these results. The level of control is on average slightly higher with the fixed BCI but the difference is not significant ($p = 0.1082, N = 20$ with two-sided, paired, Wilcoxon signed-rank test). The table shows that user satisfaction is also practically the same in both stages ($p = 0.8496, N = 20$ with two-sided, paired, Wilcoxon signed-rank test). The distribution of individual subject differences in self-assessed control and satisfaction for the two stages are not significantly different from the zero-mean normal distribution ($p = 0.1084, N = 20$ for control and $p = 0.7157, N = 20$ for satisfaction with two-sided t-tests). No differences between the two conditions are found regarding user alertness and feeling of improvement in performance (6.6 ± 2.2 vs 6.4 ± 2.0, $p = 0.3819, N = 20$ and 6.0 ± 2.1 vs 6.0 ± 1.8, $p = 0.7927, N = 20$ respectively, with two-sided, paired, Wilcoxon signed-rank tests).

As anticipated, in the scale from 1 to 10, subjects reported that the overall training paradigm exerted low physical (2.9 ± 2.0) but high mental effort (8.9 ± 1.3). The gamified approach and relatively relaxed game dynamics should probably be credited with lower than in conventional training approaches levels of frustration (4.5 ± 1.9) and time pressure perceived (4.2 ± 2.6), although these figures remain considerably high. Adverse effects like headaches were negligible (1.4 ± 0.9) whereas eye (4.1 ± 2.5) and muscle (2.6 ± 2.0) fatigue manifested in acceptable levels. The overall user comfort was marked as above average (3.6 ± 2.3, where 1 denotes a comfortable and 10 a very uncomfortable training experience), however, these results suggest that additional work is needed towards improving user experience in MI BCI training. Out of 20 participants, 15 declared they would be willing to continue training with this protocol in additional sessions.

3) EVOLUTION OF BCI PERFORMANCE

Having established the soundness of the implemented coadaptive training paradigm, Fig. 3 illustrates the results addressing the main question posed, i.e., whether adaptation affects the evolution of single-sample BCI accuracy, the main variable of interest. Fig. 3b shows that the accuracy per run achieved by subjects individually tends to fluctuate around each subject's average performance with no clear increasing or decreasing trend for the majority of participants. Consequently, the average population performance is flat and does not resemble a learning curve. The biggest difference between two consecutive runs is that between the first two runs and it is still marginal (1.8%). There is no statistical significance between accuracy distributions extracted for any pair of (consecutive or not) runs, even without correcting for multiple comparisons ($p > 0.15, N = 20$ in all cases with two-sided, paired Wilcoxon signed-rank tests). Following this outcome, no significant differences in single-sample accuracy are denoted when pulling together runs executed with (runs 1-3) and without (runs 4-5) adaptation (Online, red boxplots, Fig. 3a), despite a negligible increase in the non-adaptive condition (median of 58.7% versus 60.9%, average $61.2 \pm 7.6(s.d.)\%$ versus $60.6 \pm 7.6(s.d.)\%$, $p = 0.455, N = 20$ with two-sided, paired Wilcoxon signed-rank test). Overall, these findings determine that switching off the adaptation had no impact whatsoever (neither positive nor negative) on the subjects' BCI performance.

B. SUBJECT LEARNING

Shedding further light on potential learning outcomes, we examine complementary aspects of separability between the EEG patterns of resting and MI, namely, their "offline", cross-validated single-sample classification accuracy, the

TABLE 2. Control and user satisfaction self-assessment (scale 1-10).

Subject ID	Control		Satisfaction	
	Adaptive	Fixed	Adaptive	Fixed
S1	10	10	9	9
S2	8	9	9	9
S3	6	8	9	9
S4	8	6	10	9
S5	8	7	8	8
S6	7	6	6	7
S7	7	9	8	9
S8	7	8	10	10
S9	8	9	5	8
S10	7	6	6	7
S11	7	7	8	7
S12	4	7	4	6
S13	4	5	6	4
S14	8	8	10	10
S15	3	6	6	6
S16	3	6	5	6
S17	6	7	7	7
S18	4	1	7	7
S19	3	3	5	4
S20	5	8	9	7
Mean±s.d.	6.2±2.1	6.8±2.1	7.4±1.9	7.5±1.7

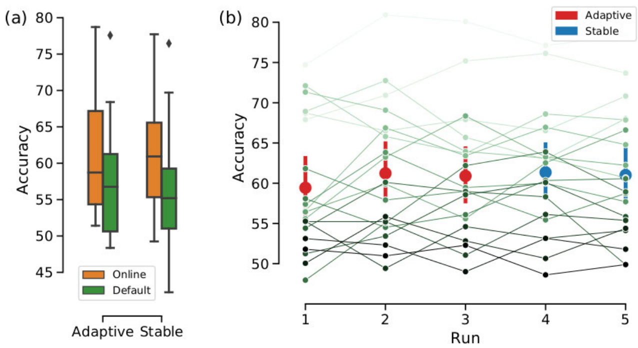

FIGURE 3. Single-sample accuracy. (a) Boxplots with median, 25th and 75th percentiles of single-sample classification accuracy per condition (Adaptive, Stable). Orange boxplots correspond to online accuracy derived with the proposed coadaptive training ("Online") and green boxplots correspond to simulated accuracy derived by applying the BCI classifier resulting from the initial 10-trial calibration to the adaptive and stable runs ("Default"). Diamonds denote statistical significance ($a = 0.99$) of the corresponding pair Online vs Default pair with two-sided, paired, Wilcoxon signed-rank tests. (b) Individual subject and average (across subjects) single-sample classification accuracy with standard deviation per run. Adaptive runs (1-3) shown in red and stable runs (4-5) in blue.

Kullback-Leibler Divergence (KLD) between these two distributions, the r^2 coefficient of determination between each subject's most discriminant SMR feature and the class labels and, lastly, the Fisher Score discriminancy (again, for each subject's most discriminant feature). Unlike online BCI classification accuracy, these metrics can assess a subject's ability to successfully modulate SMRs through MI, unconfounded by the machine learning aspects of the training paradigm (i.e., the optimality of the classifier's fit to the data), so that any improvement can be attributed to subject learning alone. Evidently, they remain good predictors of BCI performance, as revealed by the correlation with online single-sample accuracy in supplementary Fig. S2 (Offline accuracy: $r = 0.99$, $p < 10^{-17}$, KLD: $r = 0.80$, $p < 10^{-4}$, r^2: $r = 0.81$, $p < 10^{-5}$, Fisher Score: $r = 0.98$, $p < 10^{-12}$, $N = 20$ in all cases, separability values are extracted per run and then averaged over the session). Importantly, the illustrations in supplementary Fig. S2 also serve to qualitatively assess the significance of a certain decrease/increase in some separability index by revealing its approximate expected impact on classification accuracy.

Fig. 4 demonstrates that the evidence provided by these indices of subject learning converges to the fact that the subjects' MI BCI aptitude is not significantly affected ($p > 0.22, N = 20$ with two-sided, paired Wilcoxon signed-rank test for offline accuracy, r^2 and Fisher Score) by the mode of adaptation, despite a very slight increase denoted in all the metrics tested for the later, non-adaptive condition. This difference is marginally significant only in terms of KLD separability ($p = 0.025, N = 20$ with the non-parametric two-sided, paired Wilcoxon signed-rank test, $p = 0.068, N = 20$ with paired, two-sided t-test). However, this significance seems to be driven by only four "outlier responders" (S2, S4, S9, S12). Removing any two of those yields non-significant difference ($p > 0.707, N = 18$ in all cases with the non-parametric test). The absence of an important population trend comes as a result of the fact that the majority of subjects exhibit similar performances throughout the session and irrespective of the existence or not of online classifier recalibration. Furthermore, a small number of subjects who achieved better SMRs when adaptation is switched off counterbalance another as much that exhibit a performance drop. Supplementary Fig. S3 verifies that there do not seem to be clear effects of subject learning across time, by monitoring the run-wise separability of each subject and the corresponding population averages.

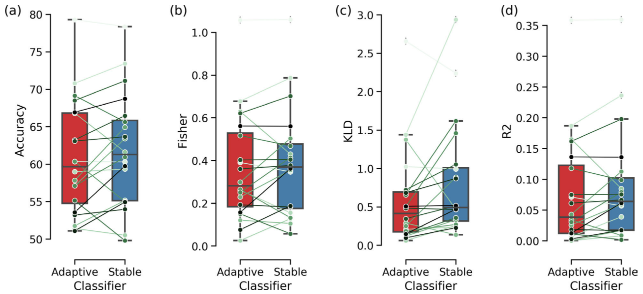

FIGURE 4. Individual subject and boxplots with median, 25^{th} and 75^{th} percentiles of separability between resting and motor imagery classes expressed as (a) offline single-sample classification accuracy, (b) Fisher Score of (assumed normal) mental class distributions for the most discriminant feature of each subject, (c) Kullback-Leibler Divergence of (assumed normal) mental class distributions in the 9-dimensional feature space used for online control and (d) r^2 coefficient of determination between the most discriminant (across the session) feature of each subject and the mental class labels. Boxplot whiskers extend to 1.5 times the respective 25^{th}-75^{th} percentile range.

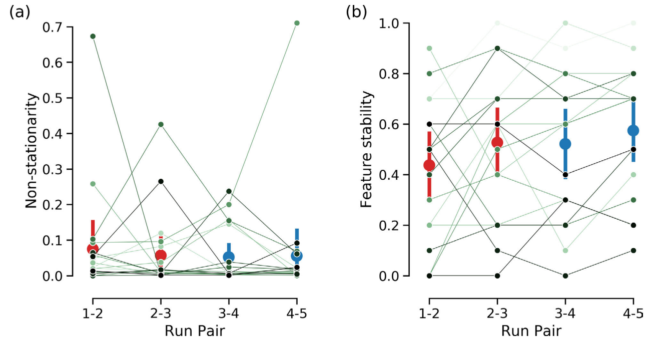

FIGURE 5. SMR pattern stability. (a) Individual subject and average (across subjects) Kullback-Leibler Divergence between the MI class distributions of each pair of consecutive runs. (b) Individual subject and average (across subjects) feature stability as the overlapping index between the sets of 10-best spatiospectral features of each pair of consecutive runs. Pairs where the second run is adaptive are shown in red, otherwise in blue.

C. STABILITY

As already noted, adaptation has been also previously implicated in potential stability loss of EEG MI correlates. To investigate this issue, we examine two different aspects of stability. First, the KLD between the MI class distributions of every pair of consecutive runs in the 9-dimensional feature space the subjects were asked to control. This metric evaluates potential non-stationarity effects during online operation. Second, by extracting offline another set of spatiospectral Power Spectral Density (PSD) features on all monitored channels with finer frequency resolution (2 Hz), we derive the overlapping index between the sets of 10-best (in terms of r^2) features in every pair of consecutive runs. This index assesses the feature stability exhibited by the participants.

Fig. 5a shows that the vast majority of subjects suffered no major non-stationarity events, with the single exception of subject S14. The population's run distributions are not statistically different for any pair of runs ($p > 0.19, N = 20$ in all cases with non-parametric two-sided, paired Wilcoxon signed-rank tests) and the average curve remains flat. Similar observations are made with regard to feature stability (Fig. 5b), despite in this case a slight, increasing tendency over the session is more evident (marginally significant, $p \sim 0.051, N = 20$ for all run pairs with two-sided, paired Wilcoxon signed-rank tests). Following the absence of strong effects, neither KLD non-stationarity of the MI feature distribution nor feature stability are significantly different when averaging runs within the two adaptation conditions (Fig. S4, $p = 0.5, N = 20$ for non-stationarity and $p = 0.16, N = 20$ for feature stability with two-sided, paired Wilcoxon signed-rank tests).

IV. DISCUSSION

This work applied a single-session, coadaptive MI BCI training protocol to a group of 20 able-bodied individuals aiming to assess potential post-adaptation effects once online BCI classifier adaptation is switched off. The main finding is that stopping machine adaptation has no significant consequences for any of the variables of interest examined, namely, BCI performance, the ability of subjects to elicit the anticipated EEG activity through MI, or the stability of SMR modulation. The evidence provided here is significant for relaxing certain concerns regarding the applicability of coadaptation as a training approach raised by several reasons in real-world scenarios where eventually or intermittently stopping decoder adaptation is either desired or imposed by the circumstances.

Our results confirm the ensemble of previous coadaptive MI BCI literature which has found that these protocols are able to bring prospective users into control of the BCI within one or a few sessions, avoiding the offline decoder calibration training stage which entails tedious and demotivating data collection [16], [27]–[30], [32]–[38]. Specifically, we have shown that the majority of participants exhibited BCI performances exceeding those of a random classifier (Table 1, Fig. 3), neurophysiologically sound EEG correlates of MI (Fig. 2, Supplementary Fig. S1) and consistent self-assessment of satisfaction and feeling of control (Table 2). The conclusions regarding BCI aptitude in the studied population do not change when considering the performances of the last adaptive run or the last run overall (Fig. 3) instead of the session-wise performance, or when testing for significant accuracy changes in the 10- and 5-lowest performers with respect to the first run. The percentage of users that did not reach adequate performance levels is also largely in agreement with previous works [27]. Establishing that the implemented protocol delivers the previously identified benefits of coadaptation has been a necessary prerequisite for studying the effects of stopping decoder adaptation, since any findings on the main question we set out to address would be meaningless out of this context.

What mainly characterizes our findings is a general lack of significant impact of training time and adaptation mode on the variables of interest. In other words, the population trends in the evolution of BCI performance, mental class separability and non-stationarity of brain patterns are not considerably affected across the executed training runs, including the transition from an adaptive to a fixed decoder. On the one hand, from a technological perspective, this should be regarded as a positive finding: as already stated, seamless BCI control and EEG SMR modulation, despite disabling decoder adaptation, enables the employment of this training paradigm in realistic BCI applications. On the other hand, some of these findings may seem to be at odds with opinions that have been popular in the field. We postulate that the mechanism of short-term coadaptation emerging from our results suggests that coadaptive training efficiently tunes the BCI model to best decode any discriminant SMR EEG activity that subjects are already able to elicit through MI at the onset of training. While the conventional open-loop calibration is also able to do that, coadaptive protocols are superior for establishing a working BCI model much faster and for engaging the user's active involvement (thus, also subject learning) up front. However, early onset of subject learning and the additional element of machine adaptivity do not imply, as proved by the results offered here, that profound subject learning (i.e., improved SMR modulation) or continuous BCI accuracy increase should be expected within such short timescales. Below, we elaborate why, in spite of dominant views, the relevant literature in fact does not contradict our conclusions.

Specifically, first, there seems to be a widespread belief that coadaptation, unlike what our results suggest (Fig. 3), should yield significant, persistent BCI performance increase. However, this expectation likely stems from over-interpretation of previous findings in this line of research. To begin with, most previous works only found significant increasing trends when isolating and testing subsets of low-performers of the studied BCI user samples [27], [29], [30], [34], [36], [37]. In the largest study [27] (and only one so far that can be said to be strongly powered) no population effect could be established and only subjects with, on average, chance performances significantly improved in the last two runs. Perdikis *et al.* [38] also observed a consistently increasing trend only for novice users within a MI BCI spelling session. Similar to our findings, Faller *et al.* [37] claim no within-session improvement in a cohort of 22 severely disabled end-users, but point out that the majority of participants were in control of the BCI by the end of the session thanks to recurrent adaptation and auto-calibration. Importantly, many works that did find performance boost thanks to coadaptation in certain user categories within a single session, obtain this outcome in a very limited number of individuals [29], [30], [34]. Furthermore, some articles reporting population effects [28] or, at least, strong individual subject effects [36] are spread over 2-3 training sessions [28], [36], [37], the extended training time being a critical factor. Hence, the existence of a general effect (i.e., anticipated for every prospective user) of coadaptive training on performance seems to still be largely debatable.

In addition to this, although we think the different methods and models that have been used in this literature should not constitute a critical factor regarding the observed general effects of coadaptive training, certain methodological choices may in fact bias the evaluation of performance evolution in a coadaptive setting. Reporting "peak" classification accuracy [36], [37] may tend to overestimate a subject's self-paced BCI aptitude and, therefore, also the assessment of performance evolution. The initial point of adaptation could also be a crucial factor. Most works employ pre-trained subject-unspecific classifiers [27]–[30], [34], [38], potentially providing a "worse" starting point in comparison to methods that employ a short subject-specific calibration period [36], [37] as also applied here. Starting from a lower level of control arguably provides a larger margin for improvement through classifier adaptation, what could manifest as larger performance increases in this type of protocols. The use of different learning rates of adaptation could also be influential to performance evolution; this aspect is very difficult to compare across the literature.

In our study, a population-wise increasing accuracy trend can be established within the first run, as implied by Fig. 3a. However, apparently, this should only reflect gradual decoder improvements as more data are received and allow to fine-tune the initial decoder (which relied on only 10 trials) to better represent the subject's EEG correlates. In other words, this trend only regards the machine learning component and should not be misinterpreted as indication of subject learning effects. Additionally, it is noted that five subjects (S7-S9, S13, S15) started with below-chance single-sample classification performance in the first run to end-up with adequate performance in the last run, showcasing fairly smooth learning curves in-between. In this subgroup, the performance increase at the end of training is marginally significant ($p = 0.0625, N = 5$ with a non-parametric two-sided, paired Wilcoxon signed-rank test, $p = 0.0022, N = 5$ with paired, two-sided, t-test). Still, given the small sample sizes and the fact that there are conventionally users who start as and remain low-performers throughout a single-session coadaptive training, we believe that these results should be interpreted with caution. In general, despite clear evidence of responders in all these works, a definite and general effect of coadaptation on BCI performance, even within low-performers or "BCI illiterate" subjects needs further corroboration.

Similarly to the overall BCI performance, no strong effect of training or adaptation mode could be found with respect to various separability metrics accompanying potential subject learning, notwithstanding a marginally significant difference between KLD before and after adaptation is switched off (Fig. 4). Of note, the latter seems to be driven by only a few subjects. This result may also be regarded as controversial given a prevalent viewpoint that coadaptation should assist subject learning [34].

Taking a critical standpoint, it must be underlined that these claims are mostly grounded on the aforementioned increase of classification accuracy often observed in coadaptation. However, this measure is a suboptimal index of subject learning [6], [40]. This is especially true in adaptive protocols where parallel, online machine learning strongly confounds classification accuracy in a way that extracted performance boost may solely be attributed to improving decoders, rather than subject learning [38]. Many works avoid reporting on the evolution of metrics defined at the feature level, as a result of which SMR modulation cannot be directly and reliably assessed [32]–[34]. Furthermore, even when such measures of subject learning are provided, the existence and magnitude of subject learning during coadaptation may remain obscure.

In coadaptive studies involving a single session, evidence of subject learning can be particularly limited. Recent articles have reported average ERD/ERS, r^2 topographic and Fisher Score maps [27], [37], but no subject learning was observed or claimed on these grounds. In some cases, evidence of improved SMR modulability is shown only for certain subjects in the studied population [30]. Another study [29] provided the most elaborate evidence of separability increase in single-session training, demonstrating r^2 improvements for the vast majority of subjects between the first and last runs. However, the magnitude and statistical significance of these improvements (i.e. their relevance for BCI performance) is not entirely clear. Another study [38] showcased negative trends of KLD separability in spite of increase in BCI and application performances, what was shown to be solely due to decoders gradually better fitting the data through adaptive BCI spelling.

In coadaptive training protocols spreading throughout 2-3 sessions, proof of subject learning has been better substantiated, though the extent of such effects remains rather inconclusive. Vidaurre *et al.* [28] show increasing mutual information that seems to correlate with increasing classification accuracy across the population, but no formal statistical assessment is attempted and these results are extracted across a longer training paradigm of three sessions. Faller *et al.* [36] presented spectra and bandpower maps where separability increase is evident for two subjects, but, again, there is no statistical assessment of subject learning throughout the population, with many subjects failing to exhibit a clear enhancement of SMR patterns through training. "Ceiling effects" cannot fully account for the absence of an overall learning effect, since many low performers remain unable to control the BCI at the end of the training, and the single-sample accuracy of even the best performers rarely approaches the saturation point (100%). Clearer evidence of SMR learning during coadaptation has been extracted in truly longitudinal training approaches [35], what has been also the case in works that applied intermittent, periodic (rather than simultaneous) decoder adaptation [6], [8], [46] or other types of self-paced BCIs [39], [45], [47].

In light of this discussion, we believe that the absence of a universally strong subject learning outcome in this work, given the short training time imposed, should be anticipated. Considering the existence of several "responders" in the recruited cohort, our findings are highly congruous with those of other coadaptive studies. The tendency for higher separability in the later, non-adaptive condition–even if not statistically significant in most cases, the slight increase in classification accuracy that is more pronounced between the first and second run, as well as the aforementioned subgroup of people that exhibited accuracy enhancement, show that, like in relevant works, also in this study some sort of learning effects did take place; however, we find it more sensible that those reflect the gradual habituation of certain subjects to the interface rather than the kind of consolidated BCI skill learning that has been shown in longitudinal MI BCI studies [6], [8], [35], [46].

Concerning the impact of adaptation on brain pattern stability, our investigation revealed no significant non-stationarity of the MI brain patterns except for a single subject (Fig. 5a) and no major changes in feature stability (Fig. 5b), although in the latter case a borderline significant trend for greater stability was observed (increase between first and last run pair, $p = 0.0506, N = 20$ with a non-parametric two-sided, paired Wilcoxon signed-rank test and $p = 0.0581, N = 20$ with paired, two-sided, t-test). These results are in line with several other studies proposing that, although non-stationarity may manifest at any time during BCI operation, the phenomenon tends to be far more frequent and intense at the transition from one session to the next [16], [28], [38]. Consequently, non-stationarity and feature stability are not significantly different between the two adaptation modes examined, either (supplementary Fig. S4). Still, the (marginally significant) higher feature stability at the end of the session may suggest that fixed decoders help subjects to produce more consistent SMRs, as has been often hypothesized [6], [40]. However, a more careful inspection reveals that this stability boost happened incrementally within both the adaptive (from the 2nd to the 3rd run) and the non-adaptive (from the 4th to the 5th run) phase of this experiment, while also the individual subject curves exhibit large variability in this respect. Hence, no safe conclusions can be reached on this point.

Summarizing the main conclusions of this discussion, we posit that the basic merit of coadaptive training is its ability to gradually (re)-calibrate on-the-fly suitable decoders able to exploit the subjects' spontaneous and pre-existing (rather than concurrently learned) SMR modulation skills. This enables users to control an MI BCI after only short training, alleviating potential major non-stationarities in the process. This result has been reproduced in all relevant studies and has rendered offline (re-)calibration redundant. The slight increase of population average BCI accuracy between the first and second run (Fig. 3b) which does not seem to be accompanied by a corresponding increase of separability in the same interval (supplementary Fig. S3) supports this claim. In other words, it seems that, in the case of the training paradigm proposed here, one adaptive run has been enough to produce a good decoder able to adequately interpret each subject's MI activity up to the extent to which the latter has been separable. On the contrary, despite frequent implications, a significant role of decoder adaptation on subject learning or on the stability of SMR patterns cannot be established, neither in this work nor in the literature, especially in a short training period, as imposed here. The absence of a profound interaction between the simultaneous machine learning and SMR modulation processes also explains the main finding of this study: stopping decoder adaptation leaves BCI performance unaffected.

The main limitation of our study is the short training time (single-session protocol). As already noted, this interval is, in all evidence, too short to allow actual subject learning effects to unfold, therefore, also the extent of the influence of adaptation on subject learning remains rather unclear. Furthermore, since non-stationarity tends to kick in mostly between BCI sessions, the full impact of adaptation on this issue cannot be studied in depth here, either. A longitudinal format should naturally follow as future work to more elaborately tap on the findings extracted here. Still, given that the main goal of implementing coadaptive training approaches is to bring people in control of the BCI the soonest possible and that single-session coadaptive paradigms have been shown to be successful in this respect, experimentally verifying post-adaptation effects after short-term coadaptation is a critical issue that had not been so far investigated.

Another problem is that the fixed order of executing the two adaptation modes confounds the experimental conditions with the training time. In other words, it cannot be safely delineated whether the observed trends of slightly increased accuracy, separability, feature stability and self-reported level of control (even though most are not statistically significant) are due to switching off the adaptation or because of the additional training effort. Clearly, the nature of our investigation (studying short-term post-adaptation effects) has dictated this order. However, cross-over designs involving several transitions between adaptive and fixed decoders could be employed in the future to disentangle the effects of these two factors.

Furthermore, although recruitment of 20 subjects is beyond the field's usual standards, the power of the study and, especially, the fact that these users are able-bodied and thus do not constitute the final end-users of BCI technology, are additional limitations that should be addressed in future work. Lastly, as far as the adaptation algorithm is concerned, the choices of using subject-unspecific bands and the focus on sustained ERD/ERS, though common in the literature, may not be optimal to maximize each user's individual performance potential.

Concluding, we have shown that switching-off the adaptation of the decoder in a single-session, MI BCI training regime has no significant impact on the subjects' BCI performance, the quality of SMR brain patterns and their stability. This finding paves the way for applying a coadaptive training methodology in real-world scenarios, where BCI adaptation may need to intermittently or definitely cease.

ACKNOWLEDGMENT

(José Diogo Cunha, Serafeim Perdikis, Sebastian Halder, and Reinhold Scherer contributed equally to this work.) Reinhold Scherer funded, Reinhold Scherer and José Diogo Cunha conceived and designed the study. José Diogo Cunha conducted the experiment. Serafeim Perdikis, José Diogo Cunha, and Sebastian Halder analyzed the results. Serafeim Perdikis, Sebastian Halder, and Reinhold Scherer wrote the paper. All authors reviewed the manuscript.

REFERENCES

[1] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, Nov. 1999.

[2] R. Scherer, G. R. Muller, C. Neuper, B. Graimann, and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: Improvement of the spelling rate," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 979–984, Jun. 2004.

[3] C. Neuper, G. Müller-Putz, R. Scherer, and G. Pfurtscheller, "Motor imagery and EEG-based control of spelling devices and neuroprostheses," *Prog. Brain Res.*, vol. 159, no. 10, pp. 393–409, 2006.

[4] 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," *J. Neural Eng.*, vol. 11, no. 3, Jun. 2014, Art. no. 036003.

[5] 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," *Artif. Intell. Med.*, vol. 59, no. 2, pp. 111–120, Oct. 2013.

[6] S. Perdikis, L. Tonin, S. Saeedi, C. Schneider, and J. D. R. Millán, "The cybathlon BCI race: Successful longitudinal mutual learning with two tetraplegic users," *PLOS Biol.*, vol. 16, no. 5, May 2018, Art. no. e2003787.

[7] R. Scherer, F. Lee, A. Schlogl, R. Leeb, H. Bischof, and G. Pfurtscheller, "Toward self-paced brain–computer communication: Navigation through virtual worlds," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 2, pp. 675–682, Feb. 2008.

[8] G. Pfurtscheller, C. Guger, G. Müller, G. Krausz, and C. Neuper, "Brain oscillations control hand orthosis in a tetraplegic," *Neurosci. Lett.*, vol. 292, no. 3, pp. 211–214, Oct. 2000.

[9] G. Pfurtscheller, G. R. Müller, J. Pfurtscheller, H. J. Gerner, and R. Rupp, "'Thought'—Control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia," *Neurosci. Lett.*, vol. 351, no. 1, pp. 33–36, Nov. 2003.

[10] R. Leeb, D. Friedman, G. R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller, "Self-paced (Asynchronous) BCI control of a wheelchair in virtual environments: A case study with a tetraplegic," *Comput. Intell. Neurosci.*, vol. 2007, pp. 1–8, Jan. 2007.

[11] T. Carlson and J. del R. Millan, "Brain-controlled wheelchairs: A robotic architecture," *IEEE Robot. Autom. Mag.*, vol. 20, no. 1, pp. 65–73, Mar. 2013.

[12] R. Leeb, L. Tonin, M. Rohm, L. Desideri, T. Carlson, and J. D. R. Millan, "Towards independence: A BCI telepresence robot for people with severe motor disabilities," *Proc. IEEE*, vol. 103, no. 6, pp. 969–982, Jun. 2015.

[13] B. J. Edelman, J. Meng, D. Suma, C. Zurn, E. Nagarajan, B. S. Baxter, C. C. Cline, and B. He, "Noninvasive neuroimaging enhances continuous neural tracking for robotic device control," *Sci. Robot.*, vol. 4, no. 31, Jun. 2019, Art. no. eaaw6844.

[14] R. Rupp, M. Rohm, M. Schneiders, A. Kreilinger, and G. R. Muller-Putz, "Functional rehabilitation of the paralyzed upper extremity after spinal cord injury by noninvasive hybrid neuroprostheses," *Proc. IEEE*, vol. 103, no. 6, pp. 954–968, Jun. 2015.

[15] A. Biasiucci, R. Leeb, I. Iturrate, S. Perdikis, A. Al-Khodairy, T. Corbet, A. Schnider, T. Schmidlin, H. Zhang, M. Bassolino, D. Viceic, P. Vuadens, A. G. Guggisberg, and J. D. R. Millán, "Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke," *Nature Commun.*, vol. 9, no. 1, pp. 1–13, Dec. 2018.

[16] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller, "Towards adaptive classification for BCI," *J. Neural Eng.*, vol. 3, no. 1, pp. R13–R23, Mar. 2006.

[17] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[18] C. Sannelli, C. Vidaurre, K.-R. Müller, and B. Blankertz, "A large scale screening study with a SMR-based BCI: Categorization of BCI users and differences in their SMR activity," *PLoS ONE*, vol. 14, no. 1, pp. 1–37, Jan. 2019.

[19] 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,'' *Artif. Intell. Med.*, vol. 59, no. 2, pp. 121–132, Oct. 2013.
[20] A. Myrden and T. Chau, ''Effects of user mental state on EEG-BCI performance,'' *Frontiers Hum. Neurosci.*, vol. 9, p. 308, Jun. 2015.
[21] C. Jeunet, E. Jahanpour, and F. Lotte, ''Why standard brain-computer interface (BCI) training protocols should be changed: An experimental study,'' *J. Neural Eng.*, vol. 13, no. 3, May 2016, Art. no. 036024.
[22] S. Rimbert, N. Gayraud, L. Bougrain, M. Clerc, and S. Fleck, ''Can a subjective questionnaire be used as brain-computer interface performance predictor?'' *Frontiers Hum. Neurosci.*, vol. 12, p. 529, Jan. 2019.
[23] A. Roc, L. Pillette, J. Mladenovic, C. Benaroch, B. N'Kaoua, C. Jeunet, and F. Lotte, ''A review of user training methods in brain computer interfaces based on mental tasks,'' *J. Neural Eng.*, vol. 18, no. 1, Feb. 2021, Art. no. 011002.
[24] C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller, ''How many people are able to operate an EEG-based brain-computer interface (BCI),'' *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 145–147, Jul. 2003.
[25] 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, Jul. 2010.
[26] M. Ahn and S. C. Jun, ''Performance variation in motor imagery brain–computer interface: A brief review,'' *J. Neurosci. Methods*, vol. 243, pp. 103–110, Mar. 2015.
[27] L. Acqualagna, L. Botrel, C. Vidaurre, A. Kübler, and B. Blankertz, ''Large-scale assessment of a fully automatic co-adaptive motor imagery-based brain computer interface,'' *PLoS ONE*, vol. 11, no. 2, pp. 1–19, Feb. 2016.
[28] C. Vidaurre, A. Schloogl, R. Cabeza, R. Scherer, and G. Pfurtscheller, ''A fully on-line adaptive BCI,'' *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1214–1219, Jun. 2006.
[29] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, ''Machine-Learning-Based coadaptive calibration for brain-computer interfaces,'' *Neural Comput.*, vol. 23, no. 3, pp. 791–816, Mar. 2011.
[30] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, ''Co-adaptive calibration to improve BCI efficiency,'' *J. Neural Eng.*, vol. 8, no. 2, Apr. 2011, Art. no. 025009.
[31] S. Perdikis, R. Leeb, and J. Del R Millan, ''Subject-oriented training for motor imagery brain-computer interfaces,'' in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 1259–1262.
[32] J. R. Wolpaw and D. J. McFarland, ''Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans,'' *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 51, pp. 17849–17854, Dec. 2004.
[33] D. J. McFarland, D. J. Krusienski, W. A. Sarnacki, and J. R. Wolpaw, ''Emulation of computer mouse control with a noninvasive brain–computer interface,'' *J. Neural Eng.*, vol. 5, no. 2, pp. 101–110, Mar. 2008.
[34] C. Vidaurre and B. Blankertz, ''Towards a cure for BCI illiteracy,'' *Brain Topography*, vol. 23, no. 2, pp. 194–198, Jun. 2010.
[35] D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw, ''Electroencephalographic (EEG) control of three-dimensional movement,'' *J. Neural Eng.*, vol. 7, no. 3, May 2010, Art. no. 036007.
[36] J. Faller, C. Vidaurre, T. Solis-Escalante, C. Neuper, and R. Scherer, ''Autocalibration and recurrent adaptation: Towards a plug and play online ERD-BCI,'' *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 3, pp. 313–319, May 2012.
[37] J. Faller, R. Scherer, U. Costa, E. Opisso, J. Medina, and G. R. Müller-Putz, ''A co-adaptive brain-computer interface for end users with severe motor impairment,'' *PLoS ONE*, vol. 9, no. 7, pp. 1–10, Jul. 2014.
[38] S. Perdikis, R. Leeb, and J. D. R. Millán, ''Context-aware adaptive spelling in motor imagery BCI,'' *J. Neural Eng.*, vol. 13, no. 3, Jun. 2016, Art. no. 036018.
[39] A. L. Orsborn, H. G. Moorman, S. A. Overduin, M. M. Shanechi, D. F. Dimitrov, and J. M. Carmena, ''Closed-loop decoder adaptation shapes neural plasticity for skillful neuroprosthetic control,'' *Neuron*, vol. 82, no. 6, pp. 1380–1393, Jun. 2014.
[40] S. Perdikis and J. D. R. Millan, ''Brain-machine interfaces: A tale of two learners,'' *IEEE Syst., Man, Cybern. Mag.*, vol. 6, no. 3, pp. 12–19, Jul. 2020.
[41] S. C. Kleih, F. Nijboer, S. Halder, and A. Kübler, ''Motivation modulates the P300 amplitude during brain–computer interface use,'' *Clin. Neurophysiol.*, vol. 121, no. 7, pp. 1023–1031, Jul. 2010.
[42] 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,'' *Biol. Psychol.*, vol. 89, no. 1, pp. 80–86, Jan. 2012.
[43] F. Lotte, F. Larrue, and C. Mühl, ''Flaws in current human training protocols for spontaneous brain-computer interfaces: Lessons learned from instructional design,'' *Frontiers Hum. Neurosci.*, vol. 7, p. 568, Sep. 2013.
[44] R. Chavarriaga, M. Fried-Oken, S. Kleih, F. Lotte, and R. Scherer, ''Heading for new shores! Overcoming pitfalls in BCI design,'' *Brain-Comput. Interface*, vol. 4, nos. 1–2, pp. 60–73, Apr. 2017.
[45] J. L. Collinger, B. Wodlinger, J. E. Downey, W. Wang, E. C. Tyler-Kabara, D. J. Weber, A. J. Mcmorland, M. Velliste, M. L. Boninger, and A. B. Schwartz, ''High-performance neuroprosthetic control by an individual with tetraplegia,'' *Lancet*, vol. 381, no. 9866, pp. 557–564, Feb. 2013.
[46] C. Neuper, G. R. Müller, A. Kübler, N. Birbaumer, and G. Pfurtscheller, ''Clinical application of an EEG-based brain–computer interface: A case study in a patient with severe motor impairment,'' *Clin. Neurophysiol.*, vol. 114, no. 3, pp. 399–409, Mar. 2003.
[47] J. M. Carmena, M. A. Lebedev, R. E. Crist, J. E. O'Doherty, D. M. Santucci, D. F. Dimitrov, P. G. Patil, C. S. Henriquez, and M. A. L. Nicolelis, ''Learning to control a Brain–Machine interface for reaching and grasping by primates,'' *PLoS Biol.*, vol. 1, no. 2, p. e42, Oct. 2003.
[48] K. Ganguly and J. M. Carmena, ''Emergence of a stable cortical map for neuroprosthetic control,'' *PLOS Biol.*, vol. 7, no. 7, pp. 1–13, Jul. 2009.
[49] J. D. Wander, T. Blakely, K. J. Miller, K. E. Weaver, L. A. Johnson, J. D. Olson, E. E. Fetz, R. P. N. Rao, and J. G. Ojemann, ''Distributed cortical adaptation during learning of a brain-computer interface task,'' *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 26, pp. 10818–10823, Jun. 2013.
[50] P. Sykacek, S. J. Roberts, and M. Stokes, ''Adaptive BCI based on variational Bayesian Kalman filtering: An empirical evaluation,'' *IEEE Trans. Biomed. Eng.*, vol. 51, no. 5, pp. 719–727, May 2004.
[51] D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw, ''Should the parameters of a BCI translation algorithm be continually adapted,'' *J. Neurosci. Methods*, vol. 199, no. 1, pp. 103–107, 2011.
[52] J. M. Carmena, ''Advances in neuroprosthetic learning and control,'' *PLoS Biol.*, vol. 11, no. 5, pp. 1–4, 2013.
[53] J. D. Cunha and R. Scherer, ''Are online co-adaptive sensorimotor rhythm brain-computer interface training paradigms effective?'' in *Proc. Int. Conf. Cyberworlds (CW)*, Oct. 2018, pp. 419–422.
[54] L. Bonnet, F. Lotte, and A. Lecuyer, ''Two brains, one game: Design and evaluation of a multiuser BCI video game based on motor imagery,'' *IEEE Trans. Comput. Intell. AI Games*, vol. 5, no. 2, pp. 185–198, Jun. 2013.
[55] Z. Wang, Y. Yu, M. Xu, Y. Liu, E. Yin, and Z. Zhou, ''Towards a hybrid BCI gaming paradigm based on motor imagery and SSVEP,'' *Int. J. Hum. Comput. Interact.*, vol. 35, no. 3, pp. 197–205, Feb. 2019.
[56] C. Neuper, R. Scherer, M. Reiner, and G. Pfurtscheller, ''Imagery of motor actions: Differential effects of kinesthetic and visual–motor mode of imagery in single-trial EEG,'' *Cognit. Brain Res.*, vol. 25, no. 3, pp. 668–677, Dec. 2005.
[57] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, ''Single-trial analysis and classification of ERP components—A tutorial,'' *NeuroImage*, vol. 56, no. 2, pp. 814–825, May 2011.
[58] G. Müller-Putz, R. Scherer, C. Brunner, R. Leeb, and G. Pfurtscheller, ''Better than random: A closer look on BCI results,'' *Int. J. Bioelectromagn.*, vol. 10, no. 1, pp. 52–55, 2008.

JOSÉ DIOGO CUNHA was born in Lisbon, Portugal, in 1995. He received the B.S. and M.S. degrees in biomedical engineering and biophysics from the Faculty of Sciences, University of Lisbon, Portugal, in 2018.

He worked as an Intern Researcher with the Institute of Neural Engineering, TU Graz, Austria, in 2018. He is currently working as a Data Scientist with Perspectum, Oxford, U.K. His research interests include the development of human–computer interaction systems making use of biosignals and gamification of medical treatment.

SERAFEIM PERDIKIS (Member, IEEE) received the M.E. degree in electrical and computer engineering from the Aristotle University of Thessaloniki (AUTH), Greece, in 2007, and the Ph.D. degree in brain-computer interaction (BCI) from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2014.

He joined the Wyss Center for Bio- and Neuro-engineering, Geneva, Switzerland, in 2015. He was a Research Engineer working on neurorehabilitation with Mindmaze SA, Switzerland, in November 2017. Since February 2019, he has been working as a Lecturer with the Computer Science and Electronic Engineering Department, University of Essex, U.K, where he is also a Researcher of the Brain-Computer Interfaces and Neural Engineering Group. His research interests include BCI, theoretical and applied machine learning, artificial intelligence, bio-signal and especially electroencephalography (EEG) and electromyography (EMG) signal processing, mutual learning and co-adaptation in BCI, as well as translational applications of BCI for assistive technology and rehabilitation.

SEBASTIAN HALDER received the degree in bioinformatics and the Ph.D. degree in brain-computer interfaces from the University of Tübingen, Tübingen, Germany, in 2006 and 2011, respectively.

Since 2011, he has been a Postdoctoral Researcher with the University of Würzburg, Würzburg, Germany, the National Rehabilitation Centre for Persons with Disabilities, Tokorozawa, Japan, and the University of Oslo, Oslo, Norway. Since January 2019, he has been an Assistant Professor/Lecturer with the Department of Computer Science and Electrical Engineering, University of Essex, Colchester, U.K. He is the author or coauthor of over 45 peer-reviewed articles. His research interests include brain–computer interfaces, measures of consciousness, and neurophysiological markers of pain.

Dr. Halder was a recipient of a fellowship awarded by the Japan Society for the Promotion of Science, in 2015, and the Alexander von Humboldt Foundation, in 2016.

REINHOLD SCHERER (Member, IEEE) received the M.S. and Ph.D. degrees in computer science from the Technische Universität Graz (TU Graz), Austria, in 2001 and 2008, respectively, and the Habilitation (venia docendi) for applied computer science, in 2016.

From 2008 to 2010, he was a Postdoctoral Researcher with the Department for Computer Science and Engineering, University of Washington, Seattle, WA, USA. From 2010 to 2016, he was an Assistant Professor. From 2011 to 2018, he was the Deputy Director of the Institute of Neural Engineering. From 2016 to 2018, he was an Associate Professor with the Institute for Neural Engineering, TU Graz. He is currently a Professor of brain-computer interfaces (BCI) and neural engineering (NE) with the University of Essex, Colchester, U.K., where he is also the Co-director of the Essex BCI-NE Laboratory. He has published many scientific articles and holds one patent. His research interests include BCI, statistical and adaptive signal processing, mobile brain and body imaging, technology-mediated rehabilitation, and assistive technologies. He is an Associate Editor of the *Scientific Reports*, *Frontiers in Neuroprosthetics*, and *Brain-Computer Interfaces* journals.