EEG non-stationarity across multiple sessions during a Motor Imagery-BCI intervention: two post stroke case series

Elaine Astrand, Jeanette Plantin, Susanne Palmcrantz and Jonatan Tidare

Abstract— **Clinical Electroencephalogram (EEG) Brain-Computer-Interface (BCI) rehabilitation largely depend on reliable information extraction from steadily evolving brain features. Non-stationary EEG feature behavior is considered a major challenge and a lot of effort has been devoted to developing adaptive methods to accommodate for this non-stationarity. However, learning- and plasticity-related mechanisms throughout a BCI intervention are additional sources of non-stationarity, that even though expected, we know very little about. In this work, we explore the evolution of Motor Imagery (MI) information extraction across multiple sessions, in two stroke patients, using a fixed and an adaptive Support Vector Machine (SVM) model. We show different behavior of the fixed SVM model for the two patients, indicating that for one patient, relevant MI-related EEG features shifted throughout the intervention. This observation calls for further investigations to better understand the evolution and shift of features across sessions, as well as the impact of using adaptive methods from a clinical outcome perspective.**

I. Introduction

Session-to-session non-stationarity of EEG signals is a well-known phenomenon that can have detrimental consequences for BCI applications. To overcome this challenge, a lot of effort has been devoted to developing machine learning methods that can be automatically adapted online to new incoming EEG data, either in a supervised e.g. [1]–[3] or in an unsupervised manner e.g. [4]–[6]. Also, so called co-adaption has been proposed that, in addition to adapting the parameters of the classification algorithm also tracks and adaptively reselects input features [7]–[9]. These methods have demonstrated successful BCI control for both healthy [7], [8] and severely motor-impaired subjects [9]. For non-clinical BCI applications, the ultimate goal is to maximize real-time BCI control and is, to a large extent, independent of the underlying brain features. However, for clinical BCI applications, the features driving the feedback need to promote clinically meaningful effects. Specifically, in a clinical BCI intervention ranging over several sessions, both desired changes in the EEG activity, related to learning and plasticity-related mechanisms [10], [11], and undesired changes, related to for example variations in the daily mental state of the user (e.g. attention [12]), recording quality [13], or small changes in electrode locations, are expected. To ensure that brain features remain meaningful, (co-) adaptive methods must be used carefully in clinical BCI interventions. As of yet, research on individual variability of session-to-session EEG non-stationarities is scarce.

In this work, we explore individual session-to-session changes of accuracy and probability score distributions of a Support Vector Machine (SVM), trained to classify Motor Imagery (MI) of hand movements versus idle in two stroke patients during an ongoing MI-BCI intervention. We specifically focus on collected data without any feedback to analyze non-stationarities without adding the complexity of feedback-induced changes in the EEG activity [14].

II. Experimental Procedures

A. Participants

Two subjects with left-side chronic ischemic stroke participated in the study (1 male and 1 female, age 34 and 71 years). Both had a poststroke duration of >6 months, and moderate arm and hand sensorimotor impairment (Fugl-Meyer Assessment for the Upper Extremity [15] = 39 and 41 points, reflex items excluded) and difficulty opening and closing the hand. None of the subjects suffered from neglect. The trial was registered at clinicaltrials.gov (NCT03994042) and complied with local rules and regulations according to the Swedish Ethics Review Committee (dnr. 2019-01577).

B. Study design

Both subjects performed a total of 16 sessions each, consisting of 3 baseline sessions followed by 12 NeuroFeedback (NF) sessions of MI NF training and one post-intervention session identical to the baseline session. During baseline sessions, EEG was simultaneously recorded while subjects performed a Motor Execution and Imagery (MEI) task without any feedback. The NF sessions consisted of two phases. First, while simultaneously recording EEG, subjects performed a short version of the MEI task without any feedback. Second, subjects performed a NF MI task with real-time continual EEG-driven feedback. The study design is

*Research supported by the Promobilia Foundation (18128), the Kamprad Family Foundation (20190119) and Eskilstuna municipality through the project Sörmlandskontraktet.

E. A. is with the School of Innovation, Design and Engineering at Mälardalen University, Högskoleplan 1, 722020 Västerås, Sweden (corresponding author: elaine.astrand@mdh.se).

J. P. is with the Department of Clinical Sciences, Danderyd Hospital, Karolinska Institute, University Department of Rehabilitation Medicine, 182 88 Stockholm, Sweden.

S.P. is with the Department of Clinical Sciences, Danderyd Hospital, Karolinska Institute, University Department of Rehabilitation Medicine, 182 88 Stockholm, Sweden.

J.T. is with the School of Innovation, Design and Engineering at Mälardalen University, Högskoleplan 1, 722020 Västerås, Sweden.

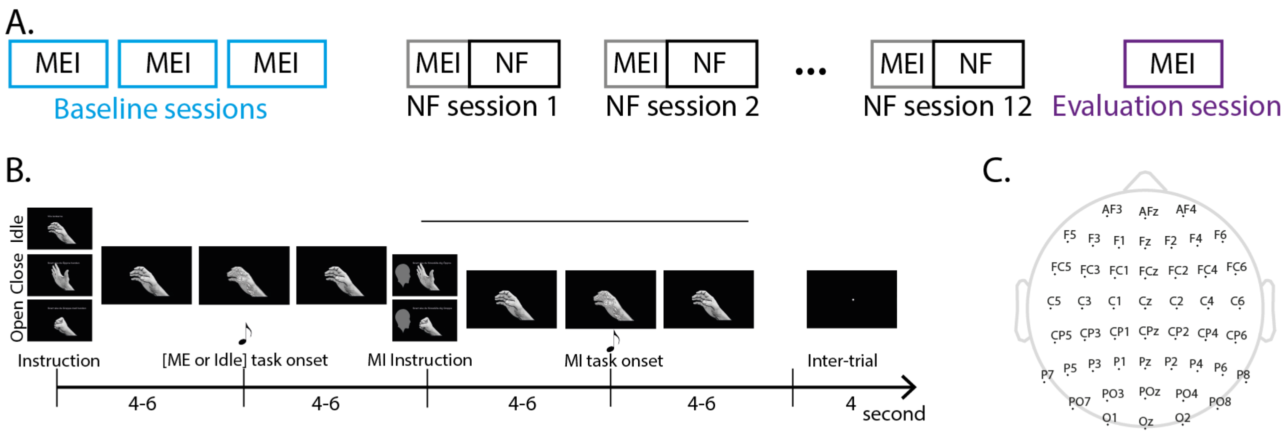

Figure 1. Experimental design. (A) The complete intervention consisted of a total of 16 sessions; 3 baseline sessions, 12 NeuroFeedback (NF) sessions and one final evaluation session identical to a baseline session. (B) Each session included a Motor Execution and Imagery (MEI) task, opening and closing hand movements, without feedback. The MEI task followed a block structure consisting of paired ME and MI tasks. An instruction (visual and auditory) to perform either open hand or close hand initiated a block. After 2.5s, the instructions were replaced by a hand in a neutral position, same for all trials. An auditory and visual cue marked the onset of a task which lasted for 4-6s. An ME task was always followed by an MI task of the same hand movement. Blocks were randomly interleaved with instructions to not do anything (idle) which followed the same trial structure as the ME trials. (C) The EEG electrodes were positioned according to the 10/10 standard electrode layout (n=48 channels).

depicted in figure 1A. For the scope of this study, only MEI task data from baseline and NF sessions are analyzed here (i.e. MEI without feedback), and in particular only MI is of interest in this study.

C. The MEI task

Seated comfortably in front of a computer screen, subjects were instructed to perform both ME and MI of their right hand (affected by the stroke) according to a task protocol displayed on the screen (detailed in figure 1B). Specifically, we asked subjects to open and close their right hand and trials of ME and MI were randomly interleaved with idling trials (in which the subjects were asked to rest their mind and not think of anything in particular). We specifically instructed the subjects to perform kinesthetic MI from a 1st person's perspective. Subjects performed a total of 120 trials (60 MI + 60 Idle), respectively, during each baseline session. During the NF sessions, subjects performed a total of 32 trials (16 MI + 16 Idle) of the MEI task.

D. EEG data acquisition

EEG recordings were carried out using 48 Ag/AgCl active electrodes (actiCHamp, Brain Products) positioned according to the international 10/10 standard (figure 1C). All impedances were kept below 30 kOhm during all sessions. The sampling rate was kept at 1 kHz. The reference was placed on the left nose wing.

E. Signal processing and feature extraction

Signal processing was performed in MATLAB (version 2019b). Raw EEG signals were first segmented into time windows of length 250 ms, and the DC component was removed from each segment. A surface Laplacian filter was further applied to reveal more localized activity (in-house scripts adapted from [16]). Time-frequency power spectral features (hereafter referred to as EEG features) were calculated using the short-time Fourier transform (windows of 250 ms), with linearly spaced frequencies from 2 Hz to 60 Hz. For further analyses, each EEG feature was averaged in a time window from 500 ms to 3000 ms after task onset. The total number of EEG features were 2832 (48 channels ×59 frequencies).

F. Classification procedure

A classification methodology was adopted to study and probe EEG features related to MI. Specifically, a Support Vector Machine (SVM) using a linear kernel [17] (MATLAB Statistics and Machine Learning toolbox) was used to discriminate EEG features related to MI (open and close hand MI trials were pooled) and idling.

Two different classification procedures were used to create and compare two SVM models; a fixed and an adaptive model. The fixed model was created by training an SVM on 131 randomly selected trials from the baseline sessions. It was then tested on data from the MEI task during each NF session (N=32). This procedure was run 100 times. The adaptive model was created by training an SVM on 100 randomly selected trials from the baseline sessions and 31 trials (N-1) from the MEI task during one NF session. The model was then tested using a Leave-One-Out (LOO) procedure on the MEI trials from the same NF session that was included in the training (N=32 trials were tested, each trial was tested using a

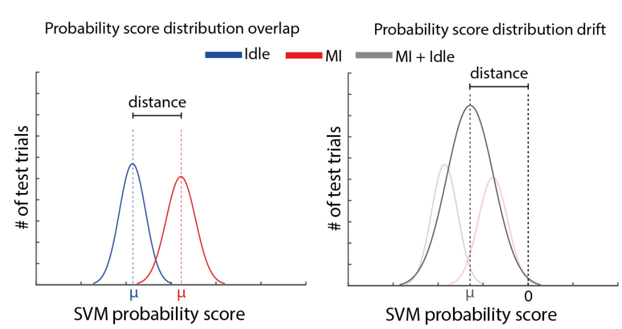

Figure 2. SVM probability score distribution (A) overlap, calculated as the difference of the mean score distributions of the two classes (MI and idle) and (B) drift, calculated as the mean of both distributions.

model trained on data that did not include the test-trial). This procedure was run 100 times. For both classification procedures, the SVM probability score distributions were analyzed in terms of overlap and drift. To calculate the overlap, we fitted the score distribution for each class (MI and Idle) with Gaussians and calculated the distance between their peaks (average($score_{MI}$)-average($score_{Idle}$)). The shorter the distance, the larger the overlap. To calculate the drift, we fitted the score distributions for both classes with one Gaussian and calculated the distance from its peak to zero (average($score_{MI+Idle}$)). These measures are illustrated in figure 2.

III. Results

We analyzed SVM accuracy and probability scores using two different calibration procedures. First, a fixed SVM model, trained on data from the MEI task during the baseline sessions, and second, an adaptive model, trained on data from the MEI tasks during both the baseline sessions and the current NF session.

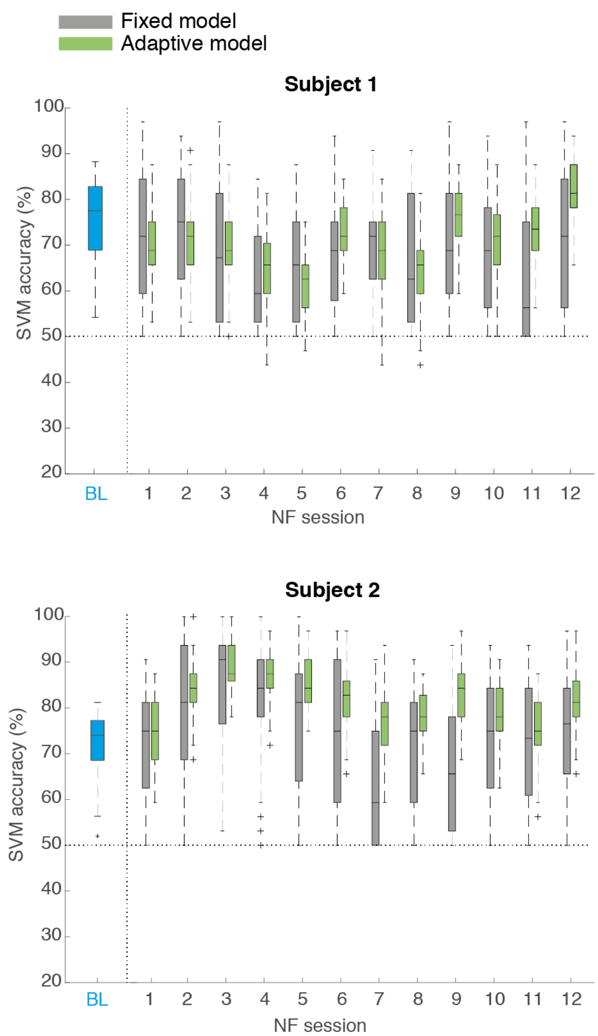

Figure 3. SVM classification accuracy during the MEI task plotted for each NF session using either a fixed model (grey) or an adaptive model (green). The accuracy during the baseline sessions is depicted by the blue box. The lines within boxes show median accuracy across 100 runs and the box edges correspond to the 25th and 75th percentiles. Outliers are indicated with crosses.

A. Fixed model: overlap and drift of SVM score distributions

First, we wanted to investigate how the fixed model managed to generalize to data recorded at different sessions. To do this, the SVM accuracy of each NF session was first analyzed using the fixed SVM model (figure 3, grey boxes). When compared to the SVM accuracy of the baseline sessions (figure 3, blue boxes), the variability of the SVM accuracy for each NF session increased, for both subjects. This might however be a reflection of the larger test-set employed for the baseline sessions (229 and 238 vs. 32 test trials). Furthermore, we can observe that the two subjects display different accuracy behavior. First, subject 1 displays accuracies with some inter-session variability, mainly varying around 68.2% (st.d. 3.3 %) with no long-term trend of either decreasing or increasing as the sessions are progressing. On the other hand, subject 2 displays increasing accuracy during the initial 3 NF sessions (BL: 72.4%+/-6.5%, NF session 3: 84.6%+/-12.6%, $p<0.01$, Wilcoxon unpaired test), followed by a gradual session-to-session decrease in accuracy to 62.8% (+/- 12.5%) on session 7. A slight increase of classification accuracy can be observed for the remaining sessions. Overall, subject 1 show an average decrease of classification accuracy across all NF sessions as compared to the BL session, while subject 2 shows an average increase (subject 1: BL: 75.8% vs. all NF sessions: 68.2%; subject 2: BL: 72.4% vs. all NF sessions: 74.3%)

By analyzing the SVM score probabilities, we can observe that a decrease in SVM accuracy could be caused by either an overlap in SVM probability score distributions (i.e. score difference around zero), a drift of the distributions or a combination of both (figure 4). Both subjects displayed score differences above zero for all sessions, although for subject 1, the score difference co-varied with the classification accuracy to a larger extent than for subject 2 (figure 5A; subject 1: $r=0.70$, $p=0.01$; subject 2: $r=0.52$, $p=0.08$, Pearson's correlation).

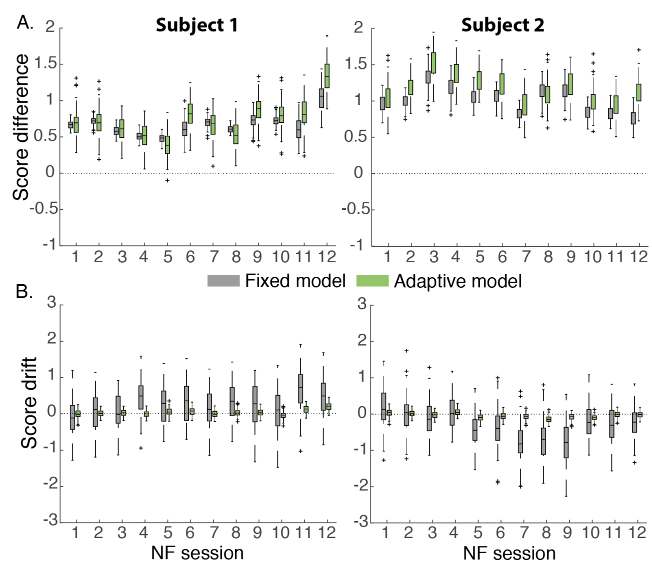

Figure 4. SVM probability score distribution (A) difference (i.e. overlap) and (B) drift during the MEI task plotted for each NF session using either a fixed model (grey) or an adaptive model (green). The lines within boxes show the median value across 100 runs and the box edges correspond to the 25th and 75th percentiles. Outliers are indicated with crosses.

Analyzing the score distribution drift reveals a small positive drift for the majority of the sessions of subject 1 while rather large negative drift for a limited number of sessions can be observed for subject 2 (figure 4B; subject 1: avg. all sessions: 0.24 +/- 0.21; subject 2: avg. sessions 5-9: -0.61 +/- 0.20). In contrast to the score difference, a drift of the score distribution correlated with classification accuracy to a higher extent for subject 2 as compared to subject 1 (figure 5B; subject 1: p=0.07; subject 2: p=0.009, Wilcoxon non-paired test).

B. Adaptive model to accommodate for both overlap and drift of score distributions

The adaptive model yielded significantly higher SVM accuracy across all sessions as compared to the fixed model only for subject 2 (figure 3, green boxes; subject 1: 68.2%+/-3.3.2 vs. 70.5+/-5.4%, p=0.002; subject 2: 74.3%+/-6.3% vs. 81.4%+/-4.5%, p=0.0005; Wilcoxon paired test).

By analyzing the score distribution difference and drift, we observe a significantly increased difference only for subject 2 (figure 4A; subject 1: 0.67 +/- 0.15 vs. 0.73 +/- 0.24, p=0.27; subject 2: 1.01 +/- 0.18 vs. 1.16 +/- 0.18, p=0.001, Wilcoxon paired test). Furthermore, the variability of score difference for each session across the 100 runs increased for both subjects when applying the adaptive model as compared to the fixed model (figure 4A; subject 1: st.d. 0.09 vs. 0.19, p=0.0005; subject 2: st.d. 0.11 vs. 0.18, p=0.0005, Wilcoxon paired test). Interestingly, the adaptive model reliably produced score distribution drifts close to zero (figure 4B; subject 1: 0.04 +/- 0.06; subject 2: -0.04 +/- 0.06).

IV. DISCUSSION

Probing the MI-related EEG features during the baseline sessions, by using a fixed SVM model, we observe different behavior of the SVM classification for the two stroke patients. Specifically, subject 2 displayed large inter-session variability of the classification accuracy, having sessions with both increased and decreased accuracy as compared to the baseline session. In contrast, the SVM accuracy of subject 1 decreased during the majority of the NF sessions as compared to the baseline session. However, it still appears to oscillate around a common mean accuracy. This behavior of subject 1 suggests that there are relatively small day-to-day changes in the EEG activity, possibly attention-related [12], but that the overall contribution of the features was not changing. From previous literature, this behavior is what one would expect from using a fixed model. In a recent report from an MI-BCI rehabilitation clinical trial with stroke patients, using a fixed Machine Learning (ML) model to control robotic online feedback, the ML accuracy varied around approximately 70% as the sessions progressed [18]. However, accuracy behavior as a function of sessions for individual subjects were not reported. We show that the SVM accuracy of subject 2 dropped steadily through NF sessions 5 to 7 and remained low for sessions 8 and 9. These results suggest that there might have been changes in the MI-related features, or alternatively subject 2 failed to perform MI. By investigating the class-wise probability score distributions, we can gain insight into the cause of the misclassifications. Specifically, NF sessions 5-9 of subject 2 show major drift in the distributions, forcing the SVM to select only one class. For some of these NF sessions the score difference was rather low, suggesting that the EEG features relevant for the fixed model could not be distinguished efficiently between MI and idle.

By analyzing the results from using the adaptive SVM, trained on data from the baseline and the current NF session, we show substantially increased SVM accuracy, for a majority of the sessions of subject 2, but not for subject 1. It has been previously shown that even simple adaptive methods can outperform static ones [2], [18], [19], and especially, adding data from the current session to update a ML model has been shown superior to adding all available data [2]. Judging from the score distribution drift, we show that the adaptive model efficiently removed any drift of the score distribution for each session. However, since subject 1 displayed less drift overall, this suggests that the impact of the adaptive model might have been reduced leading to less improvement in classification accuracy.

The interpretation that we make in terms of why the SVM accuracy dropped for some of the sessions of subject 2, is that the MI-relevant EEG features must have predominantly shifted towards other features, since the adaptive model was able to extract MI information (both by reducing score drift and increasing score difference) even when the fixed model performed poorly. We thus describe two fundamentally different non-stationarities in the EEG data, displayed by the two subjects. It has been previously described that even within a single session of training with MI feedback, controlled by using a co-adaptive approach, task-relevant features change both in space and frequency throughout the session [7], [8]. However, employing adaptive methods online in clinical MI-BCI training may have different impact for different subjects. From the observations in this work with stroke patients, subject 1 will likely train more similar EEG features throughout the intervention, while subject 2 is more likely to train different EEG features throughout the intervention. Even though the analyses in this work were performed on data collected during a MEI task without feedback, it was collected

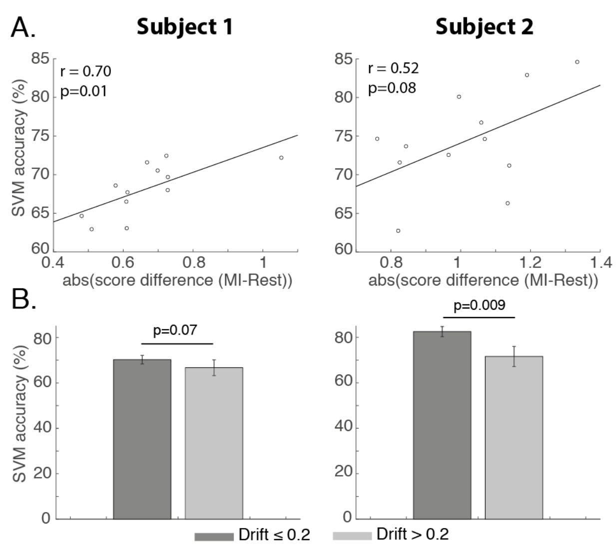

Figure 5. (A) The linear correlation between SVM accuracy and score distribution difference (i.e. overlap) across sessions for subject 1 (left) and subject 2 (right). *r* corresponds to the Pearson correlation coefficient and *p* to its corresponding significance value. (B) SVM accuracy plotted for small (≤0.2) and large (>0.2) score distribution drifts. Significance values (p) are calculated using the Wilcoxon ranksum test.

during an ongoing MI BCI-based intervention. As such, a successful intervention would be expected to, in some way alter the MI-related feature activity, the question is to what extent? To answer this question, we need to understand the neurological underpinnings and behavioral impact of feature shift during BCI training.

V. Conclusion

For both subjects, the adaptive SVM clearly demonstrated its ability to account for non-stationarities by substantially reducing any drift of the SVM score distributions. However, the adaptive model only improved the classification accuracy of subject 2. As this subject displayed large inter-session variability of the fixed model classification accuracy, enabling the adaptive model to additionally increase score differences, we believe this led to an enhanced ability to discriminate EEG data related to MI versus idle. Due to these individual differences in EEG non-stationarities across sessions, simply applying the adaptive approach in clinical BCI training after stroke to increase classification accuracy might not be optimal from a rehabilitation perspective. Future research must be devoted to developing adaptive methods to deal with undesired non-stationarity while identifying and taking advantage of desired changes in the EEG signals related to learning.

Acknowledgment

We would like to thank Hanna Bergling for helping out with the EEG setup during experiments, and Jörgen Borg and Anneli Wall for valuable discussions concerning the experimental design.

References

[1] C. Vidaurre, A. Schlogl, R. Cabeza, R. Scherer, and G. Pfurtscheller, 'Study of On-Line Adaptive Discriminant Analysis for EEG-Based Brain Computer Interfaces', *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 3, pp. 550–556, Mar. 2007

[2] D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw, 'Should the parameters of a BCI translation algorithm be continually updated?', *J Neurosci Methods*, vol. 199, no. 1, pp. 103–107, Jul. 2011

[3] A. Schlögl, C. Vidaurre, and K.-R. Müller, 'Adaptive Methods in BCI Research - An Introductory Tutorial', in *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*, B. Graimann, G. Pfurtscheller, and B. Allison, Eds. Berlin, Heidelberg: Springer, 2010, pp. 331–355.

[4] J. Blumberg, J. Rickert, S. Waldert, A. Schulze-Bonhage, A. Aertsen, and C. Mehring, 'Adaptive classification for brain computer interfaces', *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2007, pp. 2536–2539, 2007

[5] C. Vidaurre, M. Kawanabe, P. von Bünau, B. Blankertz, and K. R. Müller, 'Toward Unsupervised Adaptation of LDA for Brain–Computer Interfaces', *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 587–597, Mar. 2011

[6] B. A. S. Hasan and J. Q. Gan, 'Hangman BCI: an unsupervised adaptive self-paced Brain-Computer Interface for playing games', *Comput Biol Med*, vol. 42, no. 5, pp. 598–606, May 2012

[7] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, 'Machine-Learning-Based Coadaptive Calibration for Brain-Computer Interfaces', *Neural Computation*, vol. 23, no. 3, pp. 791–816, Dec. 2010

[8] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, 'Co-adaptive calibration to improve BCI efficiency', *J. Neural Eng.*, vol. 8, no. 2, p. 025009, Mar. 2011

[9] 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, Jul. 2014

[10] T. Wiestler and J. Diedrichsen, 'Skill learning strengthens cortical representations of motor sequences', *eLife*, vol. 2, p. e00801, Jul. 2013

[11] S. Lee, S. Ruiz, A. Caria, R. Veit, N. Birbaumer, and R. Sitaram, 'Detection of cerebral reorganization induced by real-time fMRI feedback training of insula activation: a multivariate investigation', *Neurorehabil Neural Repair*, vol. 25, no. 3, pp. 259–267, Apr. 2011

[12] C. Jeunet, B. N'Kaoua, and F. Lotte, 'Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates', *Prog Brain Res*, vol. 228, pp. 3–35, 2016

[13] E. Maby, 'Technical Requirements for High-quality EEG Acquisition', in *Brain–Computer Interfaces 2*, John Wiley & Sons, Ltd, 2016, pp. 143–161.

[14] 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. R13-23, Mar. 2006

[15] A. R. Fugl-Meyer, L. Jääskö, I. Leyman, S. Olsson, and S. Steglind, 'The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance', *Scand J Rehabil Med*, vol. 7, no. 1, pp. 13–31, 1975.

[16] M. Cohen, *Analyzing Neural Time Series Data | The MIT Press*. The MIT Press, 2014.

[17] N. Cristianini, J. Shawe-Taylor, and D. of C. S. R. H. J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[18] K. K. Ang and C. Guan, 'EEG-Based Strategies to Detect Motor Imagery for Control and Rehabilitation', *IEEE Trans Neural Syst Rehabil Eng*, vol. 25, no. 4, pp. 392–401, Apr. 2017

[19] F. Lotte *et al.*, 'A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update', *J. Neural Eng.*, vol. 15, no. 3, p. 031005, Apr. 2018