Using EEG to Distinguish between Writing and Typing for the same Cognitive Task

Xiaodong Qu[1][0000−0001−7610−6475], Qingtian Mei[1], Peiyan Liu[1], and Timothy Hickey[1]

Brandeis University, Waltham, MA 02453, USA
{xiqu,frankmei,peiyanliu,tjhickey}@brandeis.edu

Abstract. This study is designed to test the hypothesis of whether writing and typing can be detected as different patterns within the same cognitive task. We designed this pilot study with five frequently-conducted learning-related tasks. We used the four electrode Muse Headset. Sixteen healthy subjects participated in this six-session experiment. In each session, we instructed them to conduct five different one-minute tasks, including reading, copying by writing, copying by typing, answering a question by writing and answering a question by typing. We compared the performance of classifiers of different categories within the same context, including the same users, the same feature extraction approach, and the same training and testing split. Most of the machine learning and deep learning algorithms could correctly classify the five tasks (20% by chance), the best algorithms achieved an accuracy for individual subjects of up to 70% for within session training and 44% for between session training.

Keywords: Machine Learning · Deep Learning · Classification · ElectroEncephaloGraphy · EEG

1 Introduction

More and more Electroencephalography (EEG) data are available on the web, as we can search from the newly developed Google Dataset Search [16], reflecting the increased interest in EEG in both clinical and nonclinical fields, especially in emotion recognition, motor imagery, event related potential (ERP) detection, mental workload, seizure or stroke detection, Alzheimer's classification, depression, meditation and sleep analysis [8, 13, 20]. In this study we focus more on passive BCI [28], rather than the EEG controlled applications, like BCI based games [7].

Previous studies[5, 18] demonstrated that EEG signals could successfully distinguish several kinds of cognitive tasks. Such as programming in Python vs. solving Math problems; solving Math problems (GRE) vs. solving Reading problems(GRE). These experiments focused on distinguishing different cognitive tasks, but not on whether different communication modes may also have a distinguishable impact on EEG patterns. The experiment in this paper was designed to test the hypothesis of whether AI based EEG markers could distinguish both between two modes of communication: typing vs. writing, and between three cognitive states: reading vs. copying vs. answering.

Read (R) For this task, each subject was asked to read from a PDF file containing a computer science textbook on Data Structures and Algorithms. During each of the six sessions the text varied, and all of the reading material consisted mostly of text with a relatively small amount of computer code or mathematical equations.
Write Copy (WC) Each subject wrote on a blank white paper with a pen, copying the text from the textbook PDF file display on the monitor. We used the same textbook as in the "Read" task, but subjects copied different sections in each session.
Write Answer (WA) Each subject wrote a short essay using pen and paper answering the question: 'Why did you choose your major?' Although one might think that students would produce the same answer in subsequent sessions, we found that their answers in each session were substantially different.
Type Copy (TC) Each subject read a section of the same textbook PDF file, and copied what they read into an Essay Text entry box by typing on the computer keyboard. They copied different sections in each session.
Type Answer (TA) Subjects typed their answers to the question 'What is your academic plan for this semester?' The researcher asked the same question in each session, the answers varied for each session.

Fig. 1. Tasks in this experiment

In this study, each session has five tasks, we use Task Read as the baseline, selected additional four multi-label cognitive tasks that produced a text-based result, and compared the impact of communication mode (typing on a keyboard or writing with pen on paper) and cognitive task (e.g. reading or copying or composing) on the EEG signals, as shown in Figure 1, the order of the tasks are randomly shuttled in the six sessions. Our main result is that even when the tasks are designed to be similar both in communication mode and in cognitive mode, the AI based EEG markers can clearly recognize the different patterns up to a 70% level for within session training and 44% for between session training, compared with 20% by chance.

Among the frequently used portable and affordable EEG devices, [10, 15], including the Neurosky Mindwave, the Emotiv Epoc ,the Open BCI Mark IV headset, and the Muse Headband. We selected Muse headsets developed by Interaxon[21] for this study.

2 Algorithms

As many studies focus on designing new algorithms, there is also a need to analyze the existing algorithms and make full use of them. We implemented several highlighted machine learning and deep learning approaches mentioned by [8, 13, 14]. The code mentioned in this paper is available online (the Github link is hidden for the double blind review).

Linear classifiers: Both Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM) still performed fairly well after several decades since they were first implemented in this field. [3, 4, 12].

Non-linear Bayesian classifiers: Hidden Markov Models (HMM) is the algorithm we implemented. It observes a given sequence of feature vectors and outputs the probability. HMM is suitable for the classification of time series data, especially in the field of speech recognition. HMMs have performed well in several EEG signal classification studies [27].

Nearest Neighbour: k-Nearest Neighbour(kNN) [13, 14]. Usually performed from adequately to well in our previous research.

Adaboost and **Random Forest**: Boosting [9], bagging and Random Forest [6] are types of Ensemble Machine Learning Algorithms which use a group of weaker learners (e.g. random decision trees) to make a stronger classification. These have been the gold standard for many EEG classification experiments, including ours.

Transfer Learning Transfer learning has performed well with EEG data, as mentioned by [2, 13]. We used the training set with labeled data as the source domain and the testing set without labels as the target domain. We selected transductive transfer learning because the source and target tasks are the same, and the source and target domains are different but related.

CNN: Deep learning, especially Convolutional Neural Network(CNN) performed well in many previous EEG research projects [8, 11]. We implemented it with two convolution layers and one fully connected layer.

RNN, LSTM: Long Short Term Memory (LSTM), as motioned in previous research [1, 4, 8, 24], works well with time series data, we implemented it with two recurrent layers followed by two fully-connected layers.

3 Experiment

All subjects first signed an informed consent form. Then, researchers helped them to put on the Muse headbands and test the recording. The Subjects then completed an entrance survey on the computer and became familiar with the online Qualtrics system used in this experiment, especially the sample task switching notice. Next, the Official EEG recording began. A survey in Qualtrics kept track of the time and alerted the subjects to change their tasks after every 60 seconds. After subjects completed all the five tasks, the Official EEG recording stopped and subjects completed a short exit survey.

Subjects: Sixteen healthy subjects participated the experiment. Of those, data from three subjects were excluded from subsequent analysis; one for failing to participate in one of the required six sessions, and another because of considerable data loss from one of the Muse electrodes, and the third due to a very high level of noise in the electrode recordings.

Seven males and six females are included in the final data set. Ten of the retained subjects were undergraduate students, the other three were graduate students. Eight subjects were computer science majors. The average age of the subjects was 20.9.

Feature Extraction: We used the absolute Band Powers (BP) feature of the Muse headset, it is the logarithm of the power spectral density of EEG signals summed over that frequency range. [13]. The Muse headsets, are using four dry input electrodes, locations corresponded to sites TP9, AF7, AF8, and T10 [23]. The Muse EEG recording application automatically filtered out muscle artifacts, such as eye blinking. Spectral analysis was performed on-board the Muse device and then transmitted at 10 Hz to the EEG recording application on the researcher's computer. Each of these spectral snapshots consists of 20 numeric values – five spectral values for each of the four electrodes.

Data cleaning: During the EEG recording, some electrodes may have temporarily lost contact with the subjects' scalp. The result was that multiple sequential spectral snapshots from one or more electrodes had exactly the same value. When we detected this anomaly, we set that entire spectral snapshot of 20 values to 0, while keeping the time-stamped value, even if the anomaly was only detected on one of the four electrodes. Such data cleaning action resulted in a loss of 27% of the entire data.

Cross Validation: EEG data point samples, if randomly selected, could be near to each other chronologically in both the training set and the testing set. This may cause over-fitting because EEG signals changes slowly. To lessen this possible effect, we first adopted the time-wise cross validation [19, 22].

For each five minute session there are five tasks, we divided each tasks to 10 parts, evenly and contiguously, each part has 10% of the data.

Then we did a 10 fold cross validation first and realized that the first 30% of the data were predicted with low accuracy due to a task transition effect. We then cut off these transition times and only used the rest (70%) of the data. In each fold, We trained on six of the remaining seven subsets and tested on the left-out subset. Figure 2 showed a graphical representation of this seven-fold time-wise cross validation approach. The results reflect some general patterns, as shown in Figure 9.

Based on that We also did a session-wise cross validation, as shown in Figure 3, to see how the classifiers work with the data from unseen session. The results are in Figure 10.

4 Results

As shown in Figure 4, most (nine of the ten) machine learning and deep learning classifiers correctly classified the data into five classes with probability far above chance (20%). Deep learning (LSTM and CNN) and ensemble methods (Random Forest and Adaboost) outperformed other algorithms.

We first implemented the within-session approach, to investigate the general trends. Different individual algorithms generated different accuracies for each

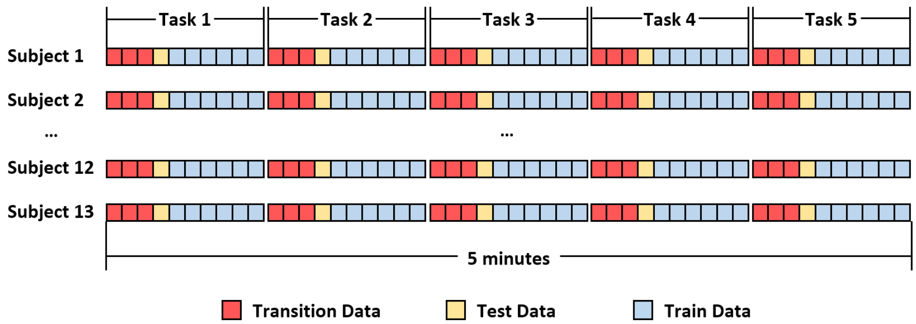

Fig. 2. Time-wise Cross Validation

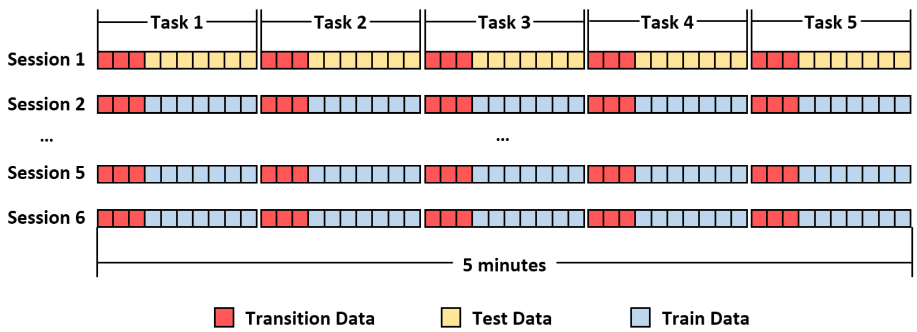

Fig. 3. Session-wise Cross Validation

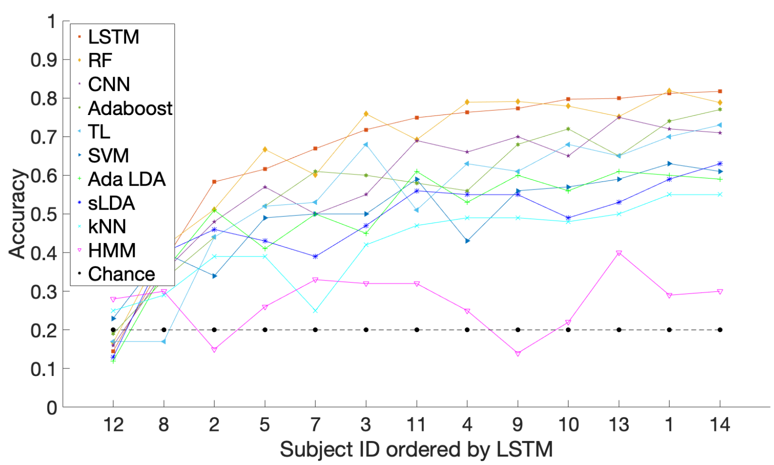

Fig. 4. Compare Algorithms, within session

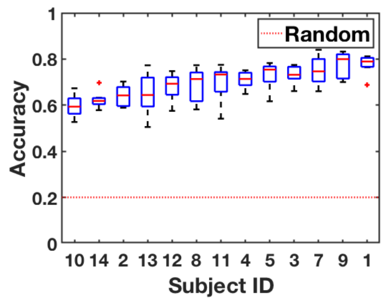

Fig. 5. Within Session, By Subject, for RF

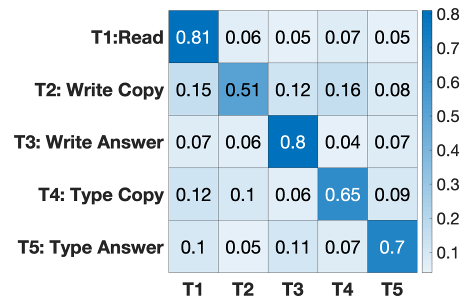

Fig. 6. Within Session, By Task, for RF

subject, such as subject 9 and subject 14, their order is different using LSTM (see Figure 4) and Random Forest (see Figure 5), but the general trends for most subjects are still similar to each other, in that some subjects data are more easily classified than others. Also we noticed a significant transition effect, as shown in Figure 9, even though the subjects physically switched to the next task quickly, the EEG signals changed slowly, in general it took 12 to 18 seconds to switch to the stable states of the next task, which corresponds to the first 20 to 30 percent of the data for that task. We then cut this first 30 percent off before running the classification algorithms again.

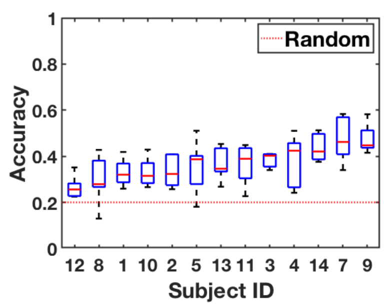

Fig. 7. Between Session, By Subject, for RF

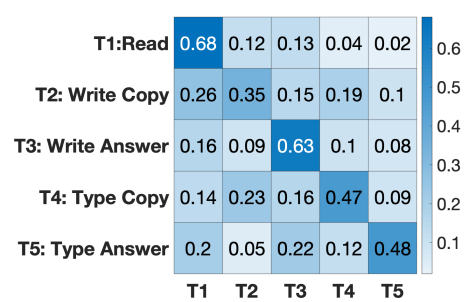

Fig. 8. Between Session, By Task, for RF

When we averaged the results by task, as shown in Figure 6, compared with 20 percent by chance, all of the five tasks are classified well with these classifiers, (here we only put RF, the other figures are available on Github).

Previous research [8, 13] has demonstrated there can be significant variations between sessions. We implemented a between session analysis, to see how well the algorithms performed when facing unseen data for the same user but on different sessions The analysis showed that even with just six sessions, the classifiers still classify with accuracies significantly above random, as show in Figure 7, and Figure 8. Task Read and Task Write Answer are still the highest, Task 2 Write Copy is still the most confusing task, which is similar to pattern in the within session approach shown in Figure 5 and Figure 6.

5 Conclusion and Discussion

Our immediate goal was to study the effectiveness of machine learning classification algorithms in distinguishing among different basic communication tasks (write vs. type) and investigate whether the method of communication (write vs. type) and the different cognitive tasks (copy or answer) could be identified. We confirmed previous findings [13, 19] that machine learning with EEG signals could discriminate two cognitive activities (copy and answer).

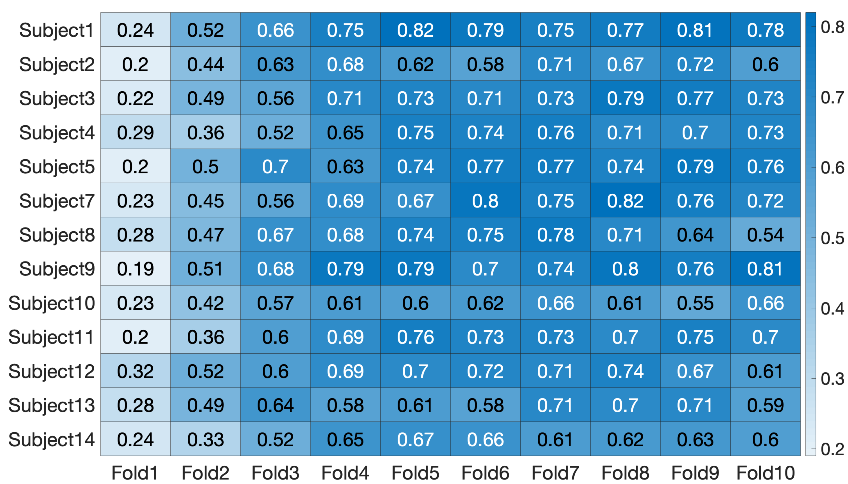

Fig. 9. Transition Within Session

We also hypothesized that this approach could reliably distinguish the two modes of communication (write and type) from one another even when the cognitive task remained the same. The confirmation of that hypothesis is a suggestion that future studies should control the communication mode variable, also this can be used for multi-label tasks in the experiment design. It is possible that the write and type tasks were distinguished by signals from the motor cortex which is located between the frontal and parietal regions of the brain.

Although these basic tasks may seem similar at the level of cortical activity, we have shown that AI classifiers can successfully capture subtle differences

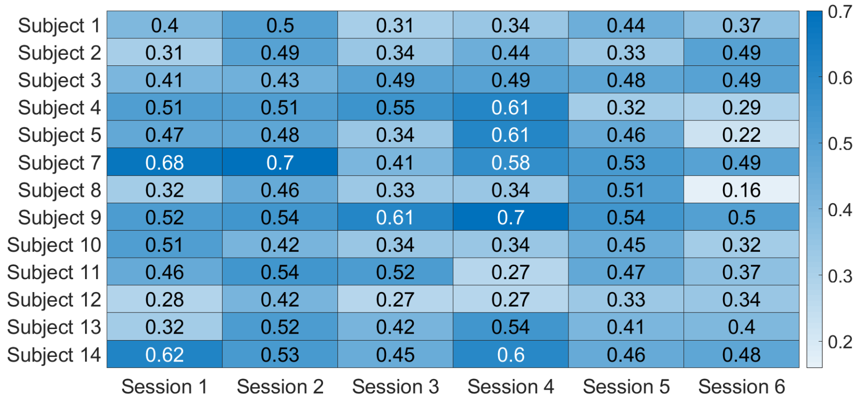

Fig. 10. Transition Between Session

between such tasks. Such result could be used in the online or offline classroom to get better understanding about teaching and learning [17, 25]. Transition effect (Figure 9) and Session variation (Figure 10) can also be sensitively detected and integrated to the future experimental design.

The main contributions of this paper are the demonstration that AI classifiers support effective approaches for classifying sets of tasks based on either copying text or answering questions using either handwriting or typing. Random forest is a fast and low-cost solution which works on most mainstream personal computers and smart phones; while LSTM is a slow and expensive solution which currently requires high-end GPUs, or the training time will jump from minutes to hours on the same personal computer compared with other classifiers like RF.

Such passive BCI approaches, as [8, 13, 28] mentioned, could be helpful for a better understanding about real-time brain signals for healthy subjects beyond medical applications[26]. More such non-clinical data sets from healthy users could in the future contribute back to the clinical research with new patterns being recognized or discovered, as well as with more effective AI classification algorithms.

Bibliography

1. Arifoglu, D., Bouchachia, A.: Activity recognition and abnormal behaviour detection with recurrent neural networks. Procedia Computer Science **110**, 86–93 (2017)
2. Arvaneh, M., Guan, C., Ang, K.K., Quek, C.: Eeg data space adaptation to reduce intersession nonstationarity in brain-computer interface. Neural computation **25**(8), 2146–2171 (2013)
3. Bashivan, P., Rish, I., Heisig, S.: Mental state recognition via wearable eeg. arXiv preprint arXiv:1602.00985 (2016)
4. Bashivan, P., Rish, I., Yeasin, M., Codella, N.: Learning representations from EEG with deep recurrent-convolutional neural networks. arXiv preprint arXiv:1511.06448 (2015)
5. Bird, J.J., Manso, L.J., Ribeiro, E.P., Ekart, A., Faria, D.R.: A study on mental state classification using eeg-based brain-machine interface. In: 2018 International Conference on Intelligent Systems (IS). pp. 795–800. IEEE (2018)
6. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
7. Coyle, D., Principe, J., Lotte, F., Nijholt, A.: Guest editorial: Brain/neuronal-computer game interfaces and interaction. IEEE Transactions on Computational Intelligence and AI in games **5**(2), 77–81 (2013)
8. Craik, A., He, Y., Contreras-Vidal, J.L.: Deep learning for electroencephalogram (eeg) classification tasks: a review. Journal of neural engineering **16**(3), 031001 (2019)
9. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences **55**(1), 119–139 (1997)
10. Gang, P., Hui, J., Stirenko, S., Gordienko, Y., Shemsedinov, T., Alienin, O., Kochura, Y., Gordienko, N., Rojbi, A., Benito, J.L., et al.: User-driven intelligent interface on the basis of multimodal augmented reality and brain-computer interaction for people with functional disabilities. In: Future of Information and Communication Conference. pp. 612–631. Springer (2018)
11. Kwak, N.S., Müller, K.R., Lee, S.W.: A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. PloS one **12**(2), e0172578 (2017)
12. Lotte, F.: Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces. Proceedings of the IEEE **103**(6), 871–890 (2015)
13. Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., Yger, F.: A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. Journal of neural engineering **15**(3), 031005 (2018)

14. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B.: A review of classification algorithms for EEG-based brain–computer interfaces. Journal of neural engineering **4**(2), R1 (2007)
15. Mihajlović, V., Grundlehner, B., Vullers, R., Penders, J.: Wearable, wireless EEG solutions in daily life applications: what are we missing? IEEE journal of biomedical and health informatics **19**(1), 6–21 (2014)
16. Noy, N.: When the web is your data lake: Creating a search engine for datasets on the web. In: Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. pp. 801–801 (2020)
17. Poulsen, A.T., Kamronn, S., Dmochowski, J., Parra, L.C., Hansen, L.K.: Eeg in the classroom: Synchronised neural recordings during video presentation. Scientific reports **7**, 43916 (2017)
18. Qu, X., Hall, M., Sun, Y., Sekuler, R., Hickey, T.J.: A personalized reading coach using wearable EEG sensors-a pilot study of brainwave learning analytics. In: CSEDU (2). pp. 501–507 (2018)
19. Qu, X., Sun, Y., Sekuler, R., Hickey, T.: EEG markers of stem learning. In: 2018 IEEE Frontiers in Education Conference (FIE). pp. 1–9. IEEE (2018)
20. Ranlund, S., Nottage, J., Shaikh, M., Dutt, A., Constante, M., Walshe, M., Hall, M.H., Friston, K., Murray, R., Bramon, E.: Resting eeg in psychosis and at-risk populations—a possible endophenotype? Schizophrenia Research **153**(1-3), 96–102 (2014)
21. Richer, R., Zhao, N., Amores, J., Eskofier, B.M., Paradiso, J.A.: Real-time mental state recognition using a wearable eeg. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 5495–5498. IEEE (2018)
22. Saeb, S., Lonini, L., Jayaraman, A., Mohr, D.C., Kording, K.P.: Voodoo machine learning for clinical predictions. Biorxiv p. 059774 (2016)
23. Seeck, M., Koessler, L., Bast, T., Leijten, F., Michel, C., Baumgartner, C., He, B., Beniczky, S.: The standardized EEG electrode array of the IFCN. Clinical neurophysiology **128**(10), 2070–2077 (2017)
24. Sha, L., Hong, P.: Neural knowledge tracing. In: International conference on brain function assessment in learning. pp. 108–117. Springer (2017)
25. Tarimo, W.T., Deeb, F.A., Hickey, T.J.: Early detection of at-risk students in cs1 using teachback/spinoza. Journal of Computing Sciences in Colleges **31**(6), 105–111 (2016)
26. Van Erp, J., Lotte, F., Tangermann, M.: Brain-computer interfaces: beyond medical applications. Computer **45**(4), 26–34 (2012)
27. Wang, M., Abdelfattah, S., Moustafa, N., Hu, J.: Deep gaussian mixture-hidden markov model for classification of eeg signals. IEEE Transactions on Emerging Topics in Computational Intelligence **2**(4), 278–287 (2018)
28. Zander, T.O., Kothe, C.: Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. Journal of neural engineering **8**(2), 025005 (2011)