A comparative study on adaptive subject-independent classification models for zero-calibration error-potential decoding

Florian M. Schönleitner[1,*], Lukas Otter[2], Stefan K. Ehrlich[1], and Gordon Cheng[1]

[1]Chair for Cognitive Systems, Department of Electrical and Computer Engineering
[2]Department of Electrical and Computer Engineering
Technische Universität München, Munich, Germany
[*]Correspondence: florian.schoenleitner@tum.de

***Abstract*— Today, a substantial part of human interaction is the engagement with artificial technological and information systems. Error-related potentials (ErrPs) provide an elegant method to improve such human-machine interaction by detecting incorrect system behaviour from the electroencephalography (EEG) signal of a human operator or user in real time. In this paper, we focus on adaptive subject-independent classification models particularly suitable for the task of ErrP decoding. As such, they provide a promising method to overcome the need of individualized decoding models, which require a time consuming calibration phase. In a comparative study we evaluate the performance of a decoding model solely trained on prior data and the effectiveness of adapting this model to a new subject. Our results show that such a generalized model can decode ErrPs with an acceptable accuracy of $(72.73 \pm 5.27)\%$ and that supervised adaptation can significantly improve the accuracy of the generalized model. Unsupervised adaptation did only prove useful for some subjects with high initial model accuracy and requires more sophisticated methods to be practical for a broader range of subjects. Consequently, our work contributes to the development of calibration-free ErrP decoding, which can potentially be used to improve human-robot interaction.**

I. Introduction

Within the last century, the ongoing development of technological and information systems fundamentally changed the way we perceive and interact with our surroundings. Nowadays, we do not only interact with other humans, but more and more frequently also with intelligent artificial systems, which can be considered as human-computer (HCI) or human-machine interaction (HMI). In most of such scenarios, the human has active control over the system and hence expects it to react by a distinct response to his command. If it does not, the system is considered flawed, which can have fatal consequences in the worst case. Consequently, one fundamental aspect of all intelligent technological systems is the correct interpretation and execution of the user's intention. Hence, a much desired ability is the evaluation of system behaviour in an online and non-interrupting manner. As error perception and evaluation are essential abilities of all humans, characteristic changes in the neurophysiological activity are measurable in the prefrontal cortex directly after observing wrong responses from an interaction partner [1], [2]. Indeed, such error-related potentials (ErrPs) provide an elegant way to detect incorrect system behaviour from the electroencephalography (EEG) signal of a human operator or user. Several groups implemented such passive brain-computer interfaces (BCIs) which monitor a subject's neurophysiological behaviour to decode ErrPs in real-time. Thus, this neuronal evaluation signal can be used to correct wrong system behaviour [3]–[9], for unsupervised adaptation [10] as well as teaching robotic skills [12], [13] and co-adapting robot behaviour during human-robot interaction [11]. Nonetheless, utilizability of ErrPs is currently impeded by the non-stationarity of the EEG signal [14], [15], which leads to variations in the signal of different subjects. As this also reflects in different feature distributions [16]–[18], simply transferring a pretrained prediction model to classify data from a new person is consequently, on the one hand, accompanied by grave performance losses [3]. On the other hand, generating subject-specific decoding models involves a time consuming initialization phase as lots of training data is required for model calibration. However, for practical subject independent systems both approaches are unpractical as the only information available would be labeled data from prior subjects and some unlabeled trials from a new subject. Fortunately, information transfer and adaptive classification models provide alternative techniques: by combining both methods it is possible to generate decoders which adapt to the subject-specific characteristics without the need for a time consuming initialization phase [10], [19]–[23].

Thus, Iturrate et al., 2011, replaced the calibration phase in ErrP decoding by inter-subject information transfer [23]. They started classification in an early stage with a classification model trained on pooled data of multiple subjects. Afterwards, they adapted the model towards a different subject by incorporating new labeled trials. This approach also proved useful for motor imagery classification. Lotte and Guan, 2009 [24] and Vidaurre et al., 2011 [20], applied a similar approach which is also feasible in unsupervised settings as no labeled information from new subjects is needed. However, their approach is not easily transferable to other paradigms as it makes some assumption about the underlying nature of inter-subject variations, which are not given by ErrPs.

We supplement the previous work by analyzing the capabilities and limitations of an adaptive classification model similar to the one used in [23] and [24] when applied to ErrP decoding. Thereby, we first test the feasibility of reusing data from other subjects and second, evaluate the effectiveness of model adaptation. In doing so, we demonstrate that model initialization from prior data combined with supervised adaptation is very effective and preferable to a minimal sample calibration. Although unsupervised adaptation did not lead to a performance increase for most subjects, it is promising especially in the case of high accuracy of the initial model.

II. Material and Methods

A. Single trial classification

Due to its simplicity and effectiveness, linear discriminant analysis (LDA) is a widely used classification method for classifying EEG signals [25]. ErrP decoding is a binary classification problem. Hence, all trials can be split into the two classes $C1$ (error), $C2$ (non-error) and the classwise Gaussian distributions with the parameters $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ as classwise means and $\boldsymbol{\Sigma}$ as the shared covariance, which is assumed to be equal for both classes. For covariance estimation, a shrinkage approach with optimal shrinkage intensity was used [26]. Single trials were classified by a LDA classifier which computes a decision boundary as follows:

$$D(\mathbf{x}) = [b, \mathbf{w}^T] \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix} \tag{1}$$

$$\mathbf{w} = \Sigma^{-1}(\boldsymbol{\mu_2} - \boldsymbol{\mu_1}) \tag{2}$$

$$b = -\mathbf{w}^T \bar{\boldsymbol{\mu}} \tag{3}$$

$$\bar{\boldsymbol{\mu}} = \frac{1}{2}(\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) \tag{4}$$

New trials are classified by computing the distance $D(\mathbf{x})$ of the feature vector to the separating hyperplane. Hence, a class decision is made based on the sign of $D(\mathbf{x})$:

$$\mathbf{x} \in \begin{cases} C_1 & \text{if } D(\mathbf{x}) < 0 \\ C_2 & \text{if } D(\mathbf{x}) > 0 \end{cases} \tag{5}$$

B. Model adaptation

For every new trial $\mathbf{x}_t$, classifier adaptation was implemented by adjusting the parameter $\boldsymbol{\mu}_i$ of the class specific probability distributions by

$$\boldsymbol{\mu}_i(t+1) = (1-\lambda)\boldsymbol{\mu}_i(t) + \frac{\lambda}{N(t)} \sum_{t=1}^{N(t)} \boldsymbol{x}_t \tag{6}$$

λ is the update coefficient which determines the rate of adaptation and $N(t)$ is the total number of available trials from the current subject.

This adaptation approach proved particularly suitable for ErrP decoding mainly because of three reasons: first, classwise mean adaptation is superior to pooled mean adaptation as previous feature analysis revealed the class dependency of variations in the individual subject's ErrPs. Similar to [23], error trials seem more variant across subjects than non-error trials. Thus, adapting only the pooled mean is not applicable as this would not account for class dependent variations. Second, adapting only the mean values but not the covariance seems sufficient as additional covariance adaptation did not lead to a noteworthy performance increase in our experiments. Further, when performing covariance adaptation $\mathbf{n(n+1)/2}$ covariance matrix entries need to be estimated from a n-sized input vector, which is - especially in the case of limited user data - highly unreliable. By restricting to mean adaptation, this ratio reduces to a linear $\mathbf{n}$ to $\mathbf{n}$ relation. Third, when comparing single trial adaptation to trial average adaptation, the latter approach has the benefit of scaling the relevance of new trials with the usage time of the system. In other words, the approach ensures convergence to the mean value of all trials of the user. The fact that this approach neglects within session adaptation is less relevant as ErrPs seem to be stable over time [27].

C. Scenarios

For a systematic evaluation of the proposed adaptive LDA model, we will compare five scenarios. These scenarios use a different amount of information from new and prior subjects.

1) 10-fold cross validation (CV): To test the performance of the proposed LDA classifier, 10-fold cross validation was performed separately for every subject. This is a standard scenario in reporting classification results in non-adaptive brain-computer interfaces and represents an optimal subject-specific model.

2) Minimal sample calibration (MSC): This scenario implements the traditional training of a subject-specific classifier with minimal samples. The calibration set consists of balanced calibration trials from the current subject only and increases over time as more trials get available. For data balancing, random undersampling was performed and the results of ten repetitions were averaged to compensate effects of differently balanced training sets. Then, the model is retrained for every new available trial of the subject. As the label information is needed for (re-)training and at least some labeled data is necessary for classifier training, this scenario is only possible in a supervised manner and cannot be used to classify the first trials of the subject. The main difference to the CV scenario is the reduced number of trials used for model training which would have a positive effect on the calibration time.

3) Generalized model (GM): A generalized model can be built without any information of the current subject by training a LDA on a balanced training set which contains all prior trials from other subjects. To compensate effects induced by random undersampling, 100 LDA models were trained on separately balanced train sets and their parameters were averaged for the final generalized model. When applied to new users, instantaneous ErrP decoding[1] is feasible,

[1] *instantaneous ErrP decoding* in this context should be understood in the sense that a meaningful classification model is available without calibration from the beginning on.

though with presumably lower classification accuracy as no individual feature characteristics are captured in the generalized model.

4) Supervised adaptation (SA): To maximize the decoding performance of the generalized model, adaptation towards a new subject was implemented according to 6. If class labels for new trials are available, supervised adaptation can be applied by using the labelled information to determine which class should be adapted.

5) Unsupervised adaptation (UA): If labelled trials are not available, only unsupervised adaptation is feasible. A label for each new trial can be predicted by the adapted generalized model $LDA_{GM}(t)$ (adapted with all available user trials at time t). Based on this label, the class parameters are adapted according to Equation (6). Provided that the adaptation rate can be initialized reasonably, this scenario would be suitable for instantaneous adaptive online ErrP decoding.

D. Experimental data

The different scenarios were evaluated on an in-house dataset[2] which contains valid[3] EEG recordings of twelve different subjects. This allowed us to simulate a system shared by different users with a sufficient amount of prior data. In a computer-based interaction scenario participants were sending directional commands depending on a visual stimulus. They perceived the movement of a cursor as feedback. To force the elicitation of error potentials in the subjects EEG, pseudo-randomized wrong cursor responses were introduced. This experimental setup is similar to those of other groups which demonstrate the feasibility of ErrP decoding [23]. EEG signals were recorded with a Brain Products actiCHamp amplifier with 32 active electrodes, arranged according to the international 10-20 system. Three channels were used for post-hoc artifact correction and all leads were referenced with respect to the average of the left and right mastoids (TP9 and TP10). With this setup, 500 trials per subject were recorded of which approximately 175 (35%) are wrong cursor responses. Data preprocessing consists of bandpass filtering, artifact correction, re-referencing, detrendering and downsampling as in [4]. Furthermore, the EEG data of each trial was segmented in a window of $[-200ms; 1000ms]$ time locked to the onset of the cursor movement. After preprocessing, temporal features were extracted similar to [4] by calculating the arithmetic mean of the signal in ten partially overlapping time windows from $[0ms; 800ms]$ after feedback presentation. In total, 270 temporal features (27 channels x 10 windows) were extracted for every trial of every subject.

E. Evaluation and reporting of results

The testing set for all scenarios consists of the last 100 trials of the current subject. As only the training but not the testing set is balanced, the performance of the different scenarios will be evaluated by comparing the balanced accuracies $bACC$ of the LDA classifiers, which is the average of the true positive and true negative rate. As the optimal adaptation parameter λ is not known a priori, we performed a grid-search by varying the adaptation rate between $[0.001; 0.05]$ in 0.001 steps. The rate which led to the highest overall $bACC$ increase for each subject and scenario was calculated post-hoc. The average of all individual optimal adaptation rates was used to initialize λ. Thus, we computed the global optimal adaptation parameter $\lambda^{SA}_{global} = 0.019$ and $\lambda^{UA}_{global} = 0.009$ for the SA and UA scenario respectively.

[2]Dataset with detailed description available at `https://github.com/stefan-ehrlich/dataset-ErrP-HRI`

[3]the data of one subject was removed from the dataset due to technical problems during data acquisition

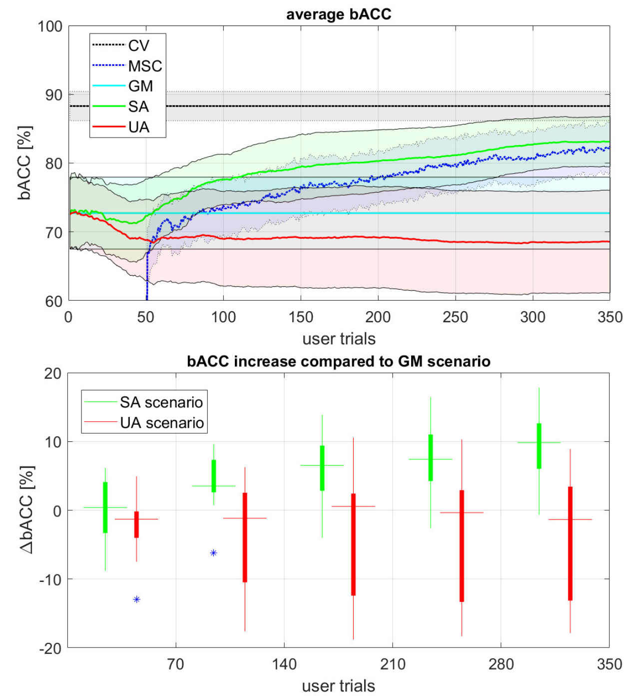

Fig. 1. Top: Averaged balanced classification accuracies for the CV (black), the MSC (blue), the GM (cyan), the SA (green) and the UA (red) scenario in a 95% confidence interval plot. Bottom: difference in model bACC due to adaptation in blocks of 70 trials; $\Delta bACC_{SA} = bACC_{SA} - bACC_{GM}$ (green) and $\Delta bACC_{UA} = bACC_{UA} - bACC_{GM}$ (red). Note: Figures are better readable when printed in color

III. Results

A. Comparing subject dependent and subject independent models

Besides the decodability of ErrPs in general, a prerequisite for calibration-free systems is the feasibility of model initialization. Here, it is practical to rely on data which is available prior to system usage as we did in the GM scenario. Comparing the averaged balanced accuracies of this scenario to the CV scenario (Figure 1, top plot), we can report three main results. First, in the case of optimal individualized decoding models reliable ErrP decoding is possible with bACC of on average $(88.29 \pm 2.13)\%$ (CV scenario). Second, reusing data of prior subjects is practicable to generate a generalized model which resulted in acceptable accuracies of $(72.73 \pm 5.27)\%$ (GM scenario). Third, as expected, there is a significant performance decrease between the accuracies of the individualized (CV scenario) and generalized (GM scenario) models of on average $-(15.56 \pm 18.65)\%$. Hence, model initialization without any data from the current subject is feasible but not optimal.

B. Effectiveness of model adaptation

To further increase the performance of the generalized model, we analyzed the effectiveness of model adaptation as described in II-B. Therefore, we defined five segments of 70 trials each in which we compared the difference between the static GM scenario and the adaptive SA and UA scenarios (Figure 1 bottom). This indicates the performance increase due to adaptation with regard to the amount of data acquired from a new subject. Looking at the adaptation performance, the beneficial effect of supervised adaptation is clearly visible: the balanced accuracies of the SA scenario (green) are significantly (two-sided Wilcoxon signed rank test, $\alpha < 0.05$) above the ones of the GM (cyan) after approximately 100 trials and also slightly above the MSC scenario (blue). During the first 50 trials adaptation could not improve the GM scenario but accuracies were approaching to $(84.23 \pm 4.02)\%$ in the later stage (Figure 1, top). Thus, a high classification accuracy only slightly below the CV scenario is established after a certain amount of trails (about 300 in our case). Afterwards, new trails are outweighed by their precursors which linearly slows down the adaptation.

Unsupervised adaptation as in the UA scenario (red curves in plot) behaves differently and leads to an unlearning effect: the accuracies drop noteworthy below the ones of the static GM scenario within the first 50 trials and converge to $(68.5 \pm 7.5)\%$, which is on average $-(3.85 \pm 5.15)\%$ below the GM scenario accuracy. Besides the pooled balanced accuracy, Figure 3 shows the bACC for the different scenarios for every individual subject. In line with Figure 1, the $bACCs$ of the supervised adaptation (green) are above the ones of the GM model for all subjects, except S_{10}. For this specific subject, adaptation is not practical as the GM model already outperforms the CV model. On the contrary, the performance of the UA scenario is not as homogeneous as the SA scenario: while unsupervised adaptation indeed improves the accuracy of the GM model for some subjects (S_3, S_5, S_6, S_8), it is detrimental for others ($S_1, S_7, S_{10}, S_{11}, S_{12}$).

As the only difference between both scenarios is the reliability of the label of new user trials - which is 100% in the SA scenario and the current model accuracy in the UA scenario - we analyzed the relation between the model accuracy and the performance increase when adapting the model. Figure 2 illustrates the performance difference of the supervised SA and unsupervised UA scenario. Depending on the current model accuracy $bACC_{model}(t)$ in the UA scenario, we computed the difference in the adaptation performance $\Delta bACC_{adapt}(t)$ to the SA scenario. One can observe that the difference between SA and UA scenario as well as the unlearning effect is especially prominent if

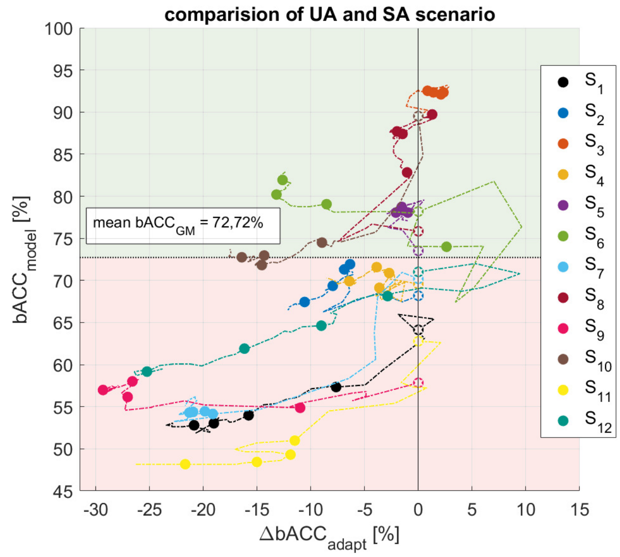

Fig. 2. Comparison of SA and UA scenario with regard to model accuracy and adaptation performance. $\Delta bACC_{adapt}(t) = bACC_{UA}(t) - bACC_{SA}(t)$ and $bACC_{model}(t)$ is the current balanced accuracy of the UA model. The dots mark the results obtained from each block of 70 trials as in Figure 1 (bottom plot). The individual initial model accuracy is equal to the GM accuracy (marked with a white star in the Plot). The plot is separated in two regions (red and green) where the interception with the y-axis is at the mean $bACC_{GM}$. Note: Figures are better readable when printed in color

the initial model accuracy is low (orange region in plot). On the contrary, unsupervised adaptation increased the GM performance for all subjects whose initial GM accuracy is above the mean GM accuracy of 72.73% (green region) and supervised adaptation is applicable[4]. Surprisingly, for one subject with very high model accuracy (S_3) unsupervised adaptation even outperformed supervised adaptation.

IV. Discussion

A. Generalized model

Within the context of inter-subject information transfer, an ideal generalized classification model should capture subject-independent feature characteristics and neglect subject-specific information which could impede transferability [19]. However, each subject's ErrPs have an individual timing, amplitude and spatial distribution. Thus, the proportion of subject-specific and therefore noisy feature information is fairly high in the generalized model. In contrast, the CV scenario is unaffected by these subject-to-subject variations which explains the higher average performances of this scenario. For the same reason we suspect that the use of simplistic classification models with a reduced number of parameters is preferable when constructing a generalized model as they are less

[4]for subject S_{10} adaptation with any of the presented methods is not helpful as the GM scenario already outperforms the CV scenario; for Subject S_6 unsupervised adaptation with the first 70 trials is superior to supervised adaptation but also not helpful

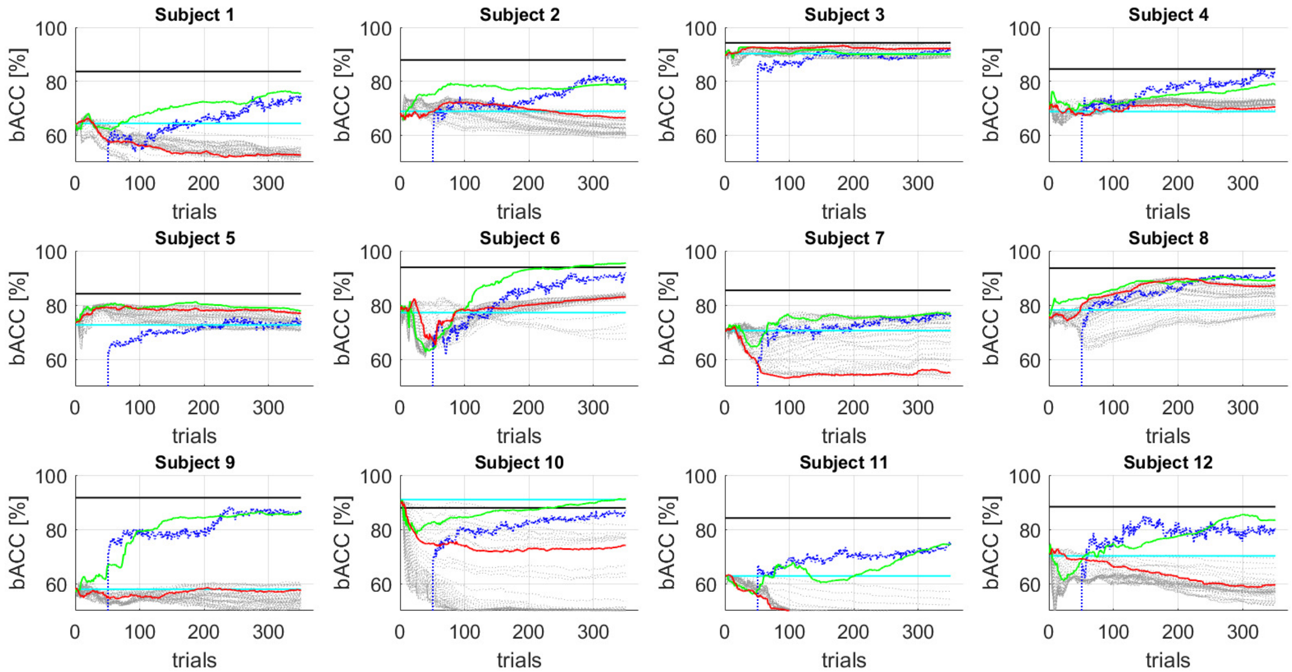

Fig. 3. Individual balanced classification accuracies for the CV (black), the MSC (blue), the GM (cyan), the SA (green) and UA scenarios (red). The grey dotted curves are the results obtained by varying the adaptation rates in the grid search approach for the UA scenario. Note: Figures are better readable when printed in color

prone to overfitting unreliable subject-independent feature characteristics from the shared training-set. Furthermore, they can be individualized efficiently as only few parameters need to be adapted. In line with [23] we also suggest that a higher number of subjects leads to a better transferability and performance of the generalized model. An increased number of subjects in the training set minimizes the risk of modeling subject-specific feature characteristics while at the same time enhances the modeling of subject-independent characteristics.

B. Adaptation

Regarding the adaptation, especially the lower performance of the UA scenario compared to its supervised counterpart is conspicuously. One obvious reason is the partially unreliable label information of new trials. While in the SA scenario, parameters of always the correct class are adapted, parameter adaptation in the UA scenario is correct only with the current accuracy of the decoder. This gives rise to the assumption of a minimal GM model accuracy needed for the UA approach to be practicable, which indeed is in line with our findings. Consequently, increasing the generalized model accuracy is of high interest as this would not only improve the classification accuracies in an early stage but most likely also the unsupervised adaptation. As we did not take into account that ErrPs of some subjects have a higher resemblance to a new subject compared to others, prior selection or prior weighting based on a suitable similarity criterion could be promising. Also the incorporation of a confidence score, which could be used to scale the adaptation rate in the unsupervised scenario, could be tested.

C. Towards plug-and-play ErrP decoding

On a broader perspective, the ultimate goal are plug-and-play systems which enable reliable subject-independent calibration-free ErrP decoding in real world environments. Here, multiple other factors which go beyond the scope of this paper have to be considered. Besides system calibration, also hardware preparation is a time consuming process in classical EEG systems which can take up to almost one hour depending on the EEG system. Here, more user friendly and simplistic systems are highly needed [28]. Furthermore, not only subject-to-subject variations but also the task dependability of ErrPs [29] limits the transferability between different settings.

V. Conclusion

This work contributes to the development of practical ErrP decoding which provides a promising method to improve human-computer and human-robot interaction by a neural evaluation signal. We demonstrated the feasibility of calibration-free systems with acceptable classification accuracies of $(72.73 \pm 5.27)\%$, which can be initialized solely from prior subjects. To compensate performance losses induced by transferring non individualized decoders, supervised adaptation towards new subject's individual feature characteristics was investigated and results were at least comparable to the minimal sample calibration scenario. Accordingly, parameter adaptation of the class-specific means of a generalized LDA model led to an average performance increase of $+(9.83 \pm 3.81)\%$ after 350 trials. Contrary, unsupervised adaptation had an unlearning effect for most subjects. Model accuracies dropped on average $-(3.85 \pm 5.15)\%$ after already 50 trials. Hence, our results demonstrate that first, instantaneous ErrP classification based on the dataset tested is feasible from the first trial on with satisfying accuracy. Second, adaptive models are already practical in supervised settings but should be applied only with particular caution in unsupervised settings.

Acknowledgement

This work was partially supported by the Elite Master Program in Neuroengineering at Technische Universität München, funded through the Elite Network Bavaria (ENB).

References

[1] M. Falkenstein, J. Hoormann, S. Christ, J. Hohnsbein, "ERP components on reaction errors and their functional significance: a tutorial", in Biological Psychology, Volume 51, Issues 23, 2000, pp. 87-107, ISSN 0301-0511.

[2] H. van Schie, R. Mars, M. Coles and H. Bekkering, "Modulation of activity in medial frontal and motor cortices during error observation", in Nature Neuroscience, vol. 7, 2004, pp. 549-554, doi: 10.1038/nn1239

[3] S. K. Ehrlich and G. Cheng, "A feasibility study for validating robot actions using eeg-based error-related potentials", International Journal of Social Robotics, pp. 1-13, Nov. 2018, issn: 1875-4805.

[4] S. Ehrlich and G. Cheng, "A neuro-based method for detecting context-dependent erroneous robot action," 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids), Cancun, 2016, pp. 477-482. doi: 10.1109/HUMANOIDS.2016.7803318

[5] M. Spüler, M. Bensch, S. Kleih, W. Rosenstiel, M. Bogdan, and A. Kübler, "Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a p300-bci", Clinical Neurophysiology, vol. 123, no. 7, pp. 1328-1337, Jul. 2012, issn: 1388-2457.

[6] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Response error correction-a demonstration of improved human-machine performance using real-time eeg monitoring", IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 11, no. 2, pp. 173-177, Jun. 2003, issn: 1534-4320.

[7] F. Kalaganis, E. Chatzilari, K. Georgiadis, S. Nikolopoulos, N. Laskaris and Y. Kompatsiaris, "An Error Aware SSVEP-based BCI," 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), Thessaloniki, 2017, pp. 775-780. doi: 10.1109/CBMS.2017.44

[8] A. Cruz, G. Pires and U. J. Nunes, "Double ErrP Detection for Automatic Error Correction in an ERP-Based BCI Speller," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 26, no. 1, pp. 26-36, Jan. 2018. doi: 10.1109/TNSRE.2017.2755018

[9] A. F. Salazar-Gomez, J. DelPreto, S. Gil, F. H. Guenther and D. Rus, "Correcting robot mistakes in real time using EEG signals," 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 2017, pp. 6570-6577. doi: 10.1109/ICRA.2017.7989777

[10] J. Blumberg, J. Rickert, S. Waldert, A. Schulze-Bonhage, A. Aertsen, and C. Mehring, "Adaptive classification for brain computer interfaces", in 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Aug. 2007, pp. 2536-2539.

[11] S. Ehrlich and G. Cheng, "Human-agent co-adaptation using error-related potentials," Journal of Neural Engineering, vol. 15, no. 6, Sep. 2018 doi: 10.1088/1741-2552/aae069

[12] I. Iturrate, J. Grizou, J. Omedes, PY. Oudeyer, M. Lopes, L. Montesano, "Exploiting Task Constraints for Self-Calibrated Brain-Machine Interface Control Using Error-Related Potentials", PLoS One. 2015;10(7):e0131491. Published 2015 Jul 1. doi:10.1371/journal.pone.0131491

[13] I. Iturrate, R. Chavarriaga, L. Montesano, J. Minguez, J. Millan, "Teaching brain-machine interfaces as an alternative paradigm to neuroprosthetics control", Scientific Reports, vol. 5, pp. 13893, Sep. 2015

[14] A. Myrden and T. Chau, "Effects of user mental state on eeg-bci performance", Frontiers in Human Neuroscience, vol. 9, p. 308, 2015, issn: 1662-5161.

[15] I. Iturrate, L. Montesano and J. Minguez, "Task-dependent signal variations in EEG error-related potentials for brain-computer interfaces," in Journal of Neural Engineering, vol. 10, no. 2, Mar. 2013, doi: 10.1088/1741-2560/10/2/026024

[16] Y. Li, H. Kambara, Y. Koike, and M. Sugiyama, "Application of covariate shift adaptation techniques in brain-computer interfaces", IEEE Transactions on Biomedical Engineering, vol. 57, no. 6, pp. 1318-1324, Jun. 2010, issn: 0018-9294.

[17] Y. Li, Y. Koike, and M. Sugiyama, "A framework of adaptive brain computer interfaces", in 2009 2nd International Conference on Biomedical Engineering and Informatics, Oct. 2009, pp. 1-5.

[18] M. Sybeldon, L. Schmit, E. Sejdic, and M. Akcakaya, "Transfer learning for eeg based bci using learn++.nse and mutual information", in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Mar. 2017, pp. 2632-2636.

[19] P. Wang, J. Lu, B. Zhang, and Z. Tang, "A review on transfer learning for brain-computer interface classification" in 2015 5th International Conference on Information Science and Technology (ICIST), Apr. 2015, pp. 315-322.

[20] 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, May 2011, issn: 0018-9294.

[21] S. Lu, C. Guan, and H. Zhang, "Learning adaptive subject-independent p300 models for eeg-based brain-computer interfaces", in 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Jun. 2008.

[22] S. Lu, C. Guan, and H. Zhang, "Unsupervised brain computer interface based on intersubject information and online adaptation", IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 17, no. 2, pp. 135-145, Apr. 2009, issn: 1534-4320.

[23] I. Iturrate, L. Montesano, R. Chavarriaga, J. del R. Millán, and J. Minguez, "Minimizing calibration time using inter-subject information of single-trial recognition of error potentials in brain-computer interfaces", in 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Aug. 2011, pp. 6369-6372.

[24] F. Lotte and C. Guan, "Learning from other subjects helps reducing Brain-Computer Interface calibration time," 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, 2010, pp. 614-617. doi: 10.1109/ICASSP.2010.5495183

[25] F. Lotte and L. Bougrain and A. Cichocki and M. Clerc and M. Congedo and A. Rakotomamonjy and F. Yger, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update", in Journal of Neural Engineering, vol. 15, no. 3, Apr. 2018, doi: 10.1088/1741-2552/aab2f2

[26] J. Schäfer and K. Strimmer, " A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics", Statistical Applications in Genetics and Molecular Biology, Nov. 2005

[27] R. Chavarriaga and J. d. R. Millan, "Learning From EEG Error-Related Potentials in Noninvasive Brain-Computer Interfaces," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 18, no. 4, pp. 381-388, Aug. 2010. doi: 10.1109/TNSRE.2010.2053387

[28] S. Ehrlich, A. Alves-Pinto, R. Lampeand and G. Cheng, "A simple and practical sensorimotor EEG device for recording in patients with special needs," in Neurotechnix2017, CogNeuroEng 2017, Symposium on Cognitive Neural Engineering, Oct. 2017.

[29] I. Iturrate, L. Montesano and J. Minguez, "Task-dependent signal variations in EEG error-related potentials for brain-computer interfaces," in Journal of Neural Engineering, vol. 10, no. 2, Mar. 2013, doi: 10.1088/1741-2560/10/2/026024