Proceedings of the
7th Graz Brain-Computer Interface Conference 2017 DOI: 10.3217/978-3-85125-533-1-77

NO TRAINING, SAME PERFORMANCE!? – A GENERIC P300 CLASSIFIER APPROACH

Andreas Pinegger[1], Gernot R. Müller-Putz[1]

[1] Institute of Neural Engineering, Graz University of Technology, Graz, Austria

E-mail: a.pinegger@tugraz.at

ABSTRACT: One of the main goals of modern brain-computer interfaces (BCIs) is that they should be simple and intuitive to use. Long-lasting training and learning periods are demotivating for the intended user. Therefore, the training should be reduced to a minimum. This particularly applies to P300-based BCIs, which are known as highly accurate and robust.

In this paper, we evaluated an approach that uses a generic classifier for P300 spelling instead of the usual personalized classifier, which users have to train before they can use the P300-based BCI. The generic classifier was calculated using the training data of 18 persons and evaluated with the data of 7 persons. Results were compared to the results achieved with personalized classifiers. We found that the generic classifier achieved comparable results regarding the effectiveness and efficiency. Therefore, our approach seems to be an appropriate, zero training alternative to personalized classifiers.

INTRODUCTION

The electroencephalogram (EEG) can be used to establish a noninvasive communication or control channel between the human brain and a computer, a so-called brain-computer interface (BCI) [1].

A very prominent BCI application is the P300 speller [2]. This type of BCI is mainly based on the positive component of an event-related potential (ERP) that appears approximately 300ms after a rare stimulus occurred among frequently occurring stimuli.

P300-based BCI provide high accuracies in combination with low illiteracy rates. Therefore, they are often used for communication and control systems. Various applications (e.g., speller [3], Brain Painting [4], music composer [5], and web browser [6]) are implemented.

Prior using such an application, training of a classifier is required. Normally, the training is performed by copy-spelling 5-10 predefined symbols and takes between 5 and 10 minutes. However, the question is, whether this training is really necessary.

Different approaches are proposed to avoid or reduce the training of the classifier. Kindermans et al. introduced a probabilistic zero training framework for ERPs [7]. They report high accuracies after a certain number of sequences. A sequence is defined as all rows and columns of the P300 matrix flashed once. However, the accuracy is still poor, when the number of sequences is limited to 3 or 4.

Lu et al. introduced a subject-independent model, learned offline from EEG of a pool of subjects, to capture common P300 characteristics [8]. They compared the learned model with a subject-specific classification model and a cross-subject model. Results indicate that this approach delivers high classification accuracies (on average approx. 84%) in combination with zero training. The number of sequences was defined with ten. No statement was given regarding the accuracies achieved with a lower number of sequences.

We asked whether the measured ERP during a P300 spelling task is stable enough to use a generic classifier. Consequently, the aim of this paper is to evaluate the power of a generic classifier (GC). The GC was calculated with the training data of eighteen P300 BCI users. The shrinkage regularized linear discriminant analysis (sLDA) was used for classification. Blankertz et al. suggested to use this method as a new standard for classifying ERPs [9]. The GC was evaluated with the data of seven users regarding the efficiency, in terms of highlighting sequences that are needed to reach certain accuracy. Effectiveness was investigated by recalculating the results of a prior study [10] with the GC: seven users had to spell four words and to control a multimedia player and a web browser with the P300 BCI. The accuracies of the online measurements and the offline simulations were compared.

MATERALS AND METHODS

Data acquisition:

The EEG data were acquired with a tap water-based biosignal amplifier (Mobita, TMSi, Oldenzaal, the Netherlands). Data were taken from six scalp electrodes (Fz, Cz, Pz, PO7, PO8, Oz) placed according to the extended international 10-20 system. A sampling rate of 250 Hz was used. The signal processing was performed in Matlab (MathWorks, Natick, USA). The EEG signal was filtered between 0.1 and 60 Hz with a 4^{th} order Butterworth band pass filter. These filter settings were chosen to compare the results of this evaluation to the results of a prior study [10].

 DOI: 10.3217/978-3-85125-533-1-77

Generic training data generation:
Eighteen healthy volunteers (5 female, mean age: 29.39, SD:12.71 years) performed a standard P300 classifier training procedure: the participants were seated in a comfortable chair approximately 60 cm away from a computer screen showing the P300 stimulation matrix, see Fig. 1. The training was performed with fifteen highlighting flashes per row and column. Each highlighting had a duration of 50 ms and the time between flashes was set to 125 ms. The task of the participants was to copy-spell five characters out of a 6 x 6 matrix filled with letters and numbers. The characters were "H3P5FU", which were equally distributed over the matrix. Elements of the matrix were highlighted with famous faces [11].

Figure 1 – P300 stimulation matrix with letters and numbers. Rows and columns were highlighted with the face of Albert Einstein.

Test data generation:
Data from the study presented in [10] were used as test data. Seven participants (1 female, mean age 25.29, SD:2.75) performed a training, hereinafter called personal training, two copy-spelling tasks, a multimedia player, and a web browser control task with the same data acquisition system, which we used to gather the training data. None of the seven participants participated in the generic training data generation measurements and the data were acquired at least half a year later than the training data. In [10] the personal training setup and signal processing were the same as described for the generic training data generation, except the word ‘’BRAIN’’ was spelled.
The copy-spelling tasks consisted of spelling 4 words with 5 letters each. The participants were advised to spell the German words ‘’SONNE’’ (engl. ‘’sun’’), ‘’BLUME’’ (engl. ‘’flower‘‘), ‘’TRAUM‘‘ (engl. ‘’dream’’), and ‘’KRAFT’’ (engl. ‘’force’’). Between the second and the third word additional tasks, see below, were performed. The users were instructed not to correct wrongly spelled letters. The matrix was the same for training and copy-spelling.
The multimedia player task was to control a multimedia player to look at pictures. The minimal number of selections was 10 and the maximum number was 15. The participants were advised to correct misclassifications. The web browser task was to look for ‘’BCI’’ in Google and to select and read the Wikipedia webpage about BCI. The minimal number of selections was 9 and the maximum number was 18. The participants were advised to correct misclassifications. The P300 matrices for the multimedia player and the web browser task were different, cf. [6].

Generic classifier creation:
The generic training data of the eighteen volunteers were divided into epochs of approximately 800 ms (204 samples) after stimulus onset. The epochs were averaged per channel and row or column. Afterwards, the data were downsampled by the factor of 12 to reduce the number of features per channel. The data of each channel were concatenated to receive one feature vector per row and column. Thus, ten target feature vectors (2 vectors * 5 characters) and fifty non-target feature vectors (10 vectors * 5 characters) were available per volunteer.
In sum, 180 target feature vectors and 900 non-target feature vectors were used to train a generic sLDA classifier.

Generic classifier evaluation:
The GC was evaluated with the test data described before. We compared the accuracies calculated with the personalized classifier (PC), i.e., the classifier trained with data from the personal training, and the GC, respectively. PC accuracies for every flashing sequence were calculated per participant by a leave-one-letter-out cross validation of the personal training data. The same personal training data were classified with the GC. Accuracies per sequence and participant were calculated to evaluate the efficiency of the GC. The efficiency is high when a small number of sequences suffice to achieve high accuracy, i.e. above 70%. This is the proposed minimal level of sufficient accuracy for BCIs, cf. [12-15].
Additionally, we compared the online accuracies of the different tasks with simulated accuracies calculated with the GC to investigate the effectiveness of the GC.

RESULTS

The spatial GC weight distribution is shown in Fig. 2. To highlight only important weights, absolute values below 0.2 are not shown.
Fig. 3 shows the average accuracies and confidence intervals of the GC and the PC using the training data of [10]. The confidence intervals show no significant differences. Interestingly, the GC on average showed better classification accuracies after sequence 13: the accuracies of the GC stayed stable at 100% or 2.9% above the PC accuracies. The proposed minimal level of

 DOI: 10.3217/978-3-85125-533-1-77

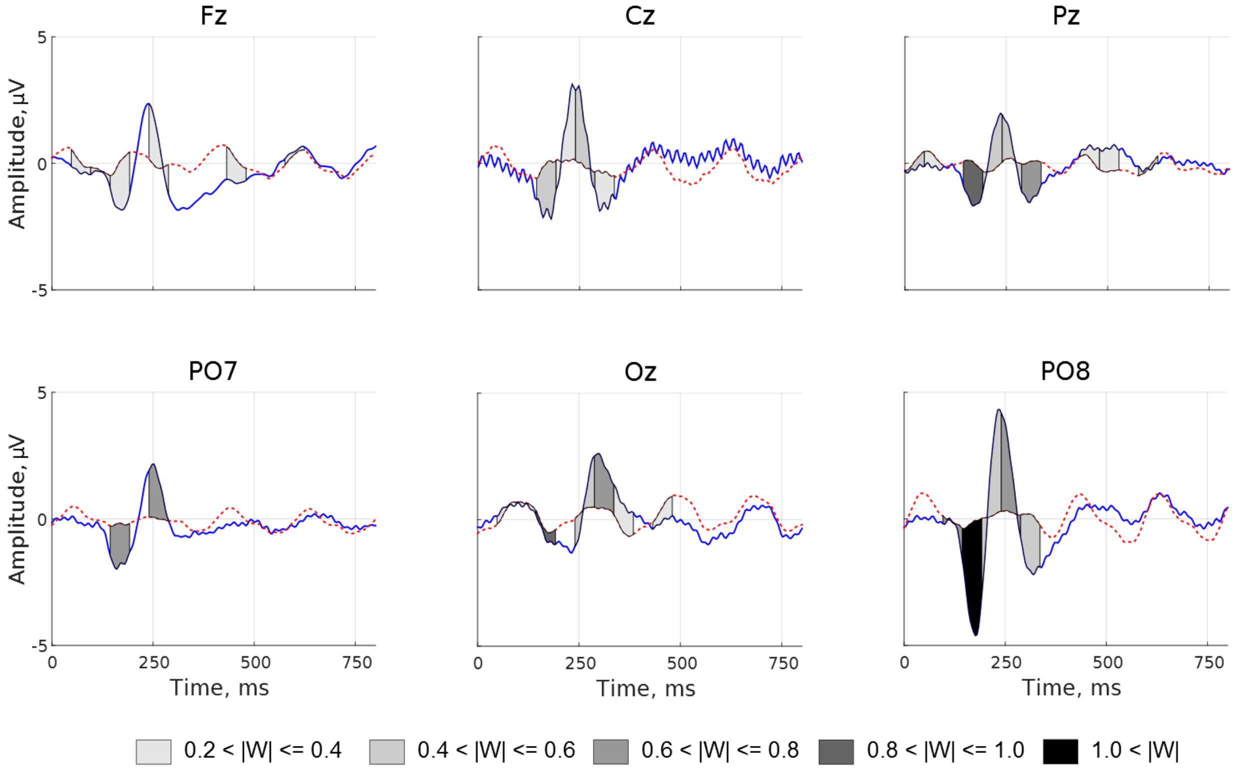

Figure 2 – The graphs show the averaged EEG data of 18 participants after targets stimulations (blue solid lines) and non-target stimulations (red dashed lines). Additionally, the weights of the GC are represented by different gray tone areas. Due to the downsampling of the signals, weights are shown as areas.

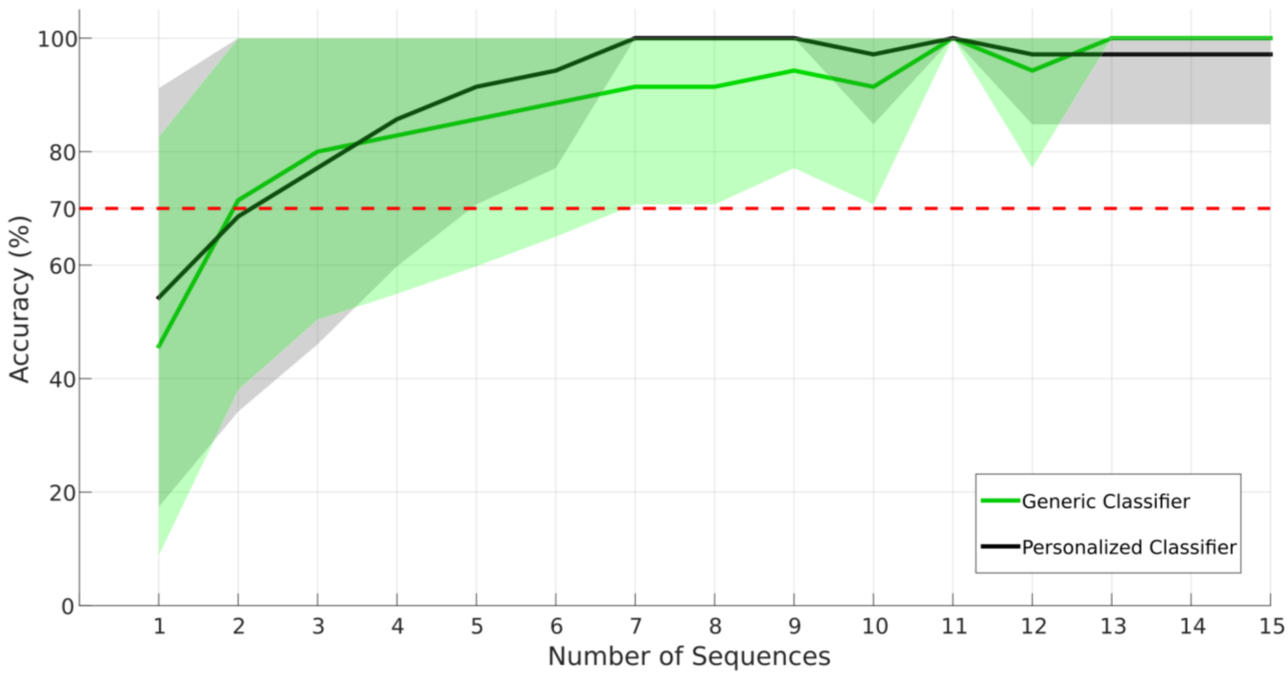

Figure 3 – Average (N=7) accuracies achieved with a certain number of sequences. The accuracies for the personal classifier were calculated with a leave-one-letter-out cross validation. Gray and green areas indicate the confidence intervals (CI) for proportions. The red dashed line indicates the minimal level of sufficient accuracy.

 DOI: 10.3217/978-3-85125-533-1-77

Table 1 – Offline (simulated) accuracies of the copy-spelling tasks using the generic classifier (GC) and the personalized classifier (PC). Different results are marked in bold. Sp1, Sp2…Spelling run 1, 2; MMP…Multimedia player; WB…Web browser.

Part.	Sequ.	GC accuracies in %						PC accuracies in %					
		Sp1	MMP	WB	Sp2	Av.	SEM	Sp1	MMP	WB	Sp2	Av.	SEM
1	8	100	100	81.8	**100**	**95.5**	10.4	100	100	**90.9**	90	95.2	10.7
2	8	100	90	**100**	80	92.5	13.2	100	**100**	90.9	**100**	**97.7**	7.5
3	9	100	100	100	100	100	0.0	100	100	100	100	100	0.0
4	10	80	91.7	88.9	80	83.9	18.4	**100**	91.7	**100**	**90**	**95.4**	10.4
5	11	70	**100**	66.7	**80**	**79.2**	20.3	**80**	64.3	**73.3**	70	71.9	22.5
6	13	100	100	**100**	100	**100**	0,0	90	100	90	100	95.0	10.9
7	14	100	100	100	**100**	**100**	0.0	100	100	100	90	97.5	7.8

sufficient accuracy (70%) was reached by the GC on average after 2 (71.4%) and by the PC after 3 (77.1%) sequences. However, the lower limits of the confidence intervals exceeded this level after 5 sequences (PC) and 7 sequences (GC), respectively, see Fig. 3.

The GC evaluation showed comparable results between the PC and GC, see Tab. 1. Differences are marked in bold. On average the GC outperformed the PC four times (range 0.3 – 7.3%) and the PC outperformed the GC two times (5.2% and 11.5%, respectively).
The average accuracies are far above the level of sufficient accuracy (70%).

DISCUSSION AND CONCLUSION

We showed that it is possible to use a P300-based BCI with zero training and high accuracies using a generic classifier. The results indicate that in terms of efficiency and effectiveness both classifiers are about equal. Moreover, the simulated GC spelling results partly outperformed the PC results.
The comparison of the accuracies for a defined number of sequences, see Fig. 3, shows that in case of a small number (between 1 and 4) no differences were detectable. For a medium number (between 5 and 10), the PC achieved better results than the GC. Finally for a large number (above 12), the GC outperformed the PC. However, the confidence intervals overlap most of the time and to make a more accurate statement more data must be taken into account.
During the spelling and control tasks the participants used a defined number of flashing sequences, see Tab. 1 second column. Comparing the averaged results indicates that participants (P2, P4) who used a medium number of sequences (between 8 and 10) would achieve better results with the PC. On the other hand, participants (P5, P6, and P7) who used a large number of sequences (above 10) would achieve higher accuracies with the GC.

One limitation of this comparison is that the presented online results were achieved with an SWLDA classifier and the simulated results were achieved with an sLDA classifier. Another limitation is that the GC was evaluated with data obtained by the same setup regarding the biosignal acquisition system, the signal processing etc. as the training data. It might be reasonably assumed that using a different biosignal acquisition system requires an adapted generic classifier.

Lu et al. also reported high P300 spelling accuracies using a generic classifier [8]. However, they performed two similar sessions with ten participants spelling the same 41 characters twice and performed a two-fold cross validation. No information was given regarding the time between the sessions and they did not evaluate the efficiency of their subject-independent model. We trained the GC with the data from different users and tasks than we evaluated it. In addition, we used different matrix sizes, cf. [6]. Finally, we used only six electrodes instead of eight in [8].

The next step would be to test the GC online with a representative number of people. In addition, it is conceivable to adapt the GC to a person by recalculating the GC with data of the actual user. Our results indicate that it should be sufficient to use a high number of sequences at the beginning to achieve almost 100% accuracy with the GC. This data can be used to recalculate the GC and adapt it to a person. Subsequently, the number of stimulation sequences can be reduced afterwards.

ACKNOWLEDGMENTS

This paper only reflects the authors' views, and funding agencies are not liable for any use that may be made of the information contained herein. The research leading to these results has received funding from the European Community's, Seventh Framework Programme FP7/2007-2013, BackHome project grant agreement number 288566.

 DOI: 10.3217/978-3-85125-533-1-77

REFERENCES

[1] J Millàn, R Rupp, GR Müller-Putz, R Murray-Smith, C Giugliemma, M Tangermann, C Vidaurre, F Cincotti, A Kübler, R Leeb, C Neuper, K-R Müller, and D Mattia. Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. Frontier in Neuroscience, vol. 4, p. 12, 2010.

[2] LA Farwell and E Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalography and Clinical Neurophysiology, vol. 70, pp. 510–523, 1988.

[3] E Donchin, KM Spencer, and R Wijesinghe. The mental prosthesis: assessing the speed of a P300-based brain-computer interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 8, pp. 174–179, 2000.

[4] J Münßinger, S Halder, S Kleih, A Furdea, V Raco, A Hösle, and A Kübler. Brain painting: first evaluation of a new brain-computer interface application with ALS-patients and healthy volunteers. Frontiers in Neuroprosthetics, vol. 4, p. 182, 2010.

[5] A Pinegger, SC Wriessnegger, and GR Müller-Putz. Sheet music by mind: Towards a brain-computer interface for composing. in Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, pp. 1053–1056, Aug 2015.

[6] S Halder, A Pinegger, I Käthner, SC Wriessnegger, J Faller, JBP Antunes, GR Müller-Putz, and A Kübler. Brain-controlled applications using dynamic P300 speller matrices. Artificial Intelligence in Medicine, vol. 63, no. 1, pp. 7 – 17, 2015.

[7] P-J Kindermans, M Tangermann, K-R Mueller, and B Schrauwen. Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP speller. Journal of Neural Engineering, vol. 11, 2014.

[8] 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, 2009.

[9] 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, pp. 814–825, 2011.

[10] A Pinegger, SC Wriessnegger, J Faller, and GR Müller-Putz. Evaluation of different EEG acquisition systems concerning their suitability for building a brain-computer interface: case studies. Frontiers in Neuroscience, 2016.

[11] T Kaufmann, SM Schulz, C Grünzinger, and A Kübler. Flashing characters with famous faces improves ERP-based brain-computer interface performance. Journal of Neural Engineering, vol. 8, no. 5, 2011.

[12] J Perelmouter, N Birbaumer. A binary spelling interface with random errors. IEEE Transactions on Rehabilitation Engineering. 2000;8:227–232.

[13] A Kübler, B Kotchoubey, J Kaiser, JR Wolpaw, N Birbaumer. Brain-computer communication: unlocking the locked in. Psychological Bulletin. 2001;127:358–375.doi:10.1037/0033-2909.127.3.358.

[14] F Nijboer, A Furdea, I Gunst, J Mellinger, DJ McFarland, N Birbaumer, et al. An auditory brain-computer interface (BCI). Journal of Neuroscience Methods. 2008;167(1):43–50. doi:10.1016/j.jneumeth.2007.02.009.

[15] A Furdea, S Halder, DJ Krusienski, D Bross, F Nijboer, N Birbaumer, et al. An auditory oddball (P300) spelling system for brain-computer interfaces. Psychophysiology. 2009;46:1–9. doi:10.1111/j.1469-8986.2008.00783.x.