Frequency Band Variations Predict EEG Single-Trial Classification Performance in Disorder of Consciousness Patients

Andrea Finke[1,3], Inga Steppacher[2,3], Johanna Kissler[2,3] and Helge Ritter[1,3]

***Abstract*— Single-trial classification of EEG data from Disorder of Consciousness patients (DoC) has proved particularly challenging. We present an approach that establishes a measure to relate the performance of single-trial classification of DoC patient EEG data with relational frequency bands and thus with their mental state. We evaluate our approach on 31 patient data sets from two studies, showing that our measure indicates for different data sets a particular likelihood for misclassifying either target or non-target class samples.**

I. INTRODUCTION

Patients with severe Traumatic Brain Injuries (TBI) often remain in a state of unresponsive wakefulness, suffering from Disorder of Consiousness (DoC). The lack of obvious, volitional reactions results in a very high number of misdiagnosis, with patients being diagnosed as vegetative when they actually are in a Minimal Conscious State (MCS). In recent years, several studies employing electroencephalography (EEG) together with Event-Related Potential (ERP) paradigms successfully investigated if DoC patients direct their attention to stimuli, e.g., [1] [2], or even understand spoken sentences, e.g., [3] [4]. Analyzing the time-locked neural responses to stimuli, the ERPs, allows to assess a patient's reaction based solely on her neural signatures, even in the complete absence of other obvious responses. One particularly successful ERP in this regard is the P300, which is a positive deflection of the neural signal amplitude after approximately 300 ms. The P300 is a response to a rare or novel stimulus, an *oddball*, in a sequence of frequent background stimuli. This EEG component is also well established in the field of Brain-Computer Interfaces (BCI), where it is typically used with visual stimuli. The best known such system is the P300 Speller [5]. Given the success of studies showing that DoC patients do indeed respond to oddball stimuli, establishing communication with DoC patients by means of a P300-based BCI becomes a possibility. However, approaches to transferring BCI technology to DOC patients have to date yielded only moderate results, e.g., [6] [7] .

In this paper, we present an approach to classify DoC patients' single-trial EEG epochs recorded during an oddball paradigm. The single-trial classification approach builds on our previous work on P300-based BCI systems. We hypothesize that the patients' mental state is not at all times identically receptive for the stimulation, which may easily lead to poor classification performance. Therefore, beyond doing single-trial classification, we analyze the relative EEG frequency bands and relate them to the classification outcome [8]. By so doing, we aim at establishing a relation between high or low degrees of variations in frequency bands as an indicator for changes in the patients' mental state and single-trial classification results. Also, frequency bands were shown to be a predictor for long term outcome in DoC patients [8].

*This work was supported by the German Ministry for Education and Research (BMBF), grant No. 16SV7789K

[1]A. Finke and H. Ritter are with the Neuroinformatics Group, Faculty of Technology, Bielefeld University, 33619 Bielefeld, Germany. {afinke,helge} at techfak.uni-bielefeld.de

[2] I. Steppacher and J. Kissler are with the Affective Neuropsychology Group, Faculty of Psychology, Bielefeld University.{inga.steppacher,johanna.kissler} at uni-bielefeld.de

[3]A. Finke, I. Steppacher, J. Kissler and H. Ritter are with the Cluster of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University.

II. DATA SETS

We report results obtained on data sets from two studies with DoC patients and one study with age-matched healthy controls. Please refer to Table I. Studies one and two did not include the same patients.

A. Experimental and EEG Setup

The experimental paradigm was a three-group auditory oddball paradigm with frequent (1000 Hz sine) and rare (1500 Hz sine) tones interspersed with novel sounds. A tone lasted for 100 ms and the interstimulus interval (ISI) was 1000 ms. One recording block consisted of a total of 500 frequent, 100 rare, and 100 novel stimuli presented in a randomized sequence. In the second study, the full recording consisted of two such stimulation blocks, approximately 0.5 hours apart. The patients were verbally instructed to silently count the high (rare) tones. EEG was recorded from 32 BioSemi (www.biosemi.com) active electrodes and sampled at 2048 Hz. The data was later re-referenced offline to the mastoids and bandpass filtered using cut-off frequencies of 0.1 Hz and 30 Hz. The procedure for the control group, all neurologically healthy, was identical, and conducted in our lab.

III. METHODS

A. Artifact Rejection

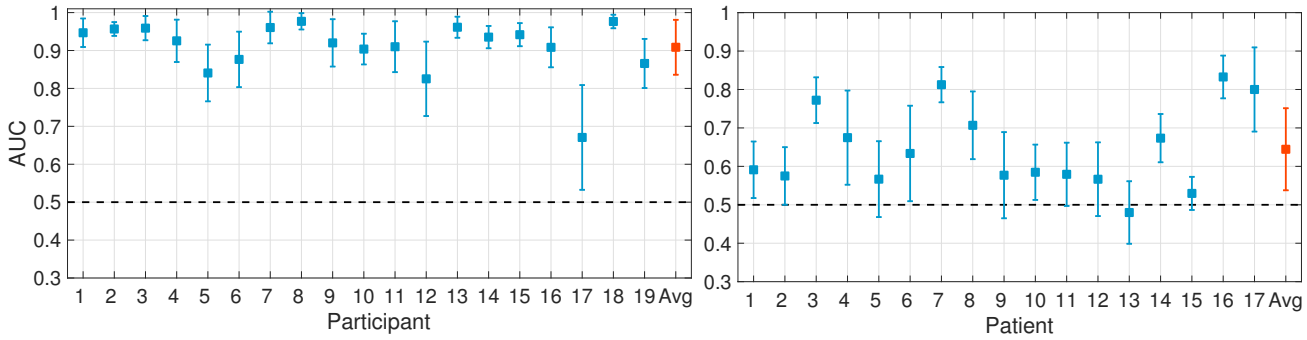
EEG data of DoC patients is especially affected by high-amplitude artifacts from eye or muscle movements. We removed artifacts using Independent Component Analysis (ICA): Non-neural sources were identified based on their spatial distribution (scalp maps) and spectral properties before back-projecting the data into sensor space.

B. Feature Extraction and Classification

For the single-trial classification, we ran pre-tests to identify the best discriminable classes for a binary classification problem, resulting in the 'novels' class to be the target

TABLE I: Basic parameters of the used data sets.

Group	No.	Age Range	Mean Age	Sex	Avg. time since injury	Range
Patients DoC 1	17	26-56	45.7	6f/11m	7.5 years	2-15 years
Patients DoC 2	14	22-67	44.3	7f / 7m	8.0 years	2-17 years
Healthy Controls (HC)	19	18-57	34.0	15f/ 4m	n/a	n/a



(a) Healthy controls

(b) DoC1 patients

Fig. 1: Results of single-trial classification of data of 17 DoC patients and a control group with 19 healthy participants.

class and the 'frequents' the non-target class. We segmented the continuous EEG data to generate epochs of 1000 ms length (2048 samples per channel). For feature extraction, we compared two approaches. The first one is based on using Principle Component Analysis (PCA) for dimension reduction by projecting the data in a lower-dimensional subspace, an approach we have previously used successfully in other studies[9] [10]. We also tried a simplified procedure where dimension reduction is achieved by downsampling the data by a factor N, that is, simply picking only the n'th sample. We chose N=50, resulting in feature vectors with a dimensionality of 1310. Binary classification was done with Fisher's Linear Discriminant Analysis (FDA). We compute the covariance matrices during FDA optimization with a shrinkage approach [11] [12].

C. Cross-Validation

We performed two types of cross-validation, a ten-fold approach to assess the overall performance of our approach on the different data sets, and a leave-one-out approach to obtain the epoch-by-epoch labels for evaluating the classification performance with respect to variations in the frequency bands. In the former, all feature vectors in a data set were first randomly shuffled to avoid one fold containing only epochs in strict temporal order. In the leave-one-out approach, all epochs except one served as the training set, and only the remaining epoch was tested, iterating until each epoch had been tested. DoC2 data, comprising two stimulation blocks, were further tested as one compound set, but also separately for each block. This approach should reveal structural variations in the neural data, resulting from the longer time span between the two stimulation blocks and possibly a patient's altered mental state, i.e., his receptiveness for the stimulation.

D. Frequency Band Variations

We computed the relative, time-varying frequency bands for all data sets based on a Short-Time Fourier Transform (STFT). We used a Tukey window function and a window size of 5 sec with an overlap of 2.5 sec. The window was shifted over the continuous data set to obtain a full assessment of the frequency variations over the whole recording. This window size was a trade-off between the assumption that relevant, mental state related shifts in the signal's frequency composition do not occur on a very small time scale but vary slowly over time, and a reasonable temporal resolution needed to assign frequency band windows to single epochs used in the classification. To obtain values for a whole frequency band, the power spectral density values at each time and frequency step were averaged over the coefficients of a whole band. These were defined according to the meaningful bands typically used in EEG research: delta (1-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta1 (13-19 Hz) and beta2 (20-29 Hz). By so doing, changes in a relative band over time are understood as this particular band gaining or losing influence with respect to the overall composition of frequencies. This approach is particularly useful to assess changes in a person's mental state, which can be inferred from this composition.

We were especially interested if the patient's EEG exhibits significant variations over time and if and how their presence or absence influences the single-trial classification of stimulus-locked epochs. To this end, we defined a measure for presence or absence of variations. A a baseline, we computed for each of the five frequency bands the average over time. Then, we set a threshold of ± 1.5 times the standard deviation over time. Frequency band windows that exceeded this threshold were marked as periods with a high variation in that band. Mean and standard deviation and thus the thresholds were assessed individually for each data set. For further analysis, all bands and windows were flagged based on whether the threshold was exceeded in positive (1) or negative (-1) direction, or the value stayed within the limit (0).

E. Evaluating Classifier Performance with Frequency Band Variations

To align the classification results with the variations in the frequency bands, we first assign to each epoch its corresponding frequency band variation window. Thus, each epoch also receives the window's flag (1,-1 or 0) explained above. Then, we use the leave-one-out cross-validation procedure to obtain a predicted label for each epoch in a data set. The predicted labels are split in two sets: The first set contains labels referring to windows with high frequency band variations (flagged -1 or 1), and the second set labels referring to windows with low variations (flagged 0). For both sets, we obtain the respective confusion matrices (i.e., True Positives (TP), False Negatives (FN), True Negatives (TN), or False Positives (FP)) using the ground truth. We create a statistics by accumulating the TP, FN, TN, and FP in each set (obtaining eight values overall). These values are normalized by the number of all target and non-target epochs, respectively. Finally, we compute the ratio of *correctly* classified epochs assigned to a high variation window and to *misclassified* epochs assigned to such a window. We will denote this ratio by *RCE*. RCE is <1 when the proportion of misclassified epochs is *higher* in phases with a high variation in a frequency band than the proportion of correctly classified epochs, and >1 when this relation is inverted. An RCE <1 thus indicates a higher likelihood of misclassifying epochs during phases of high frequency band variation.

IV. RESULTS

We first report the classification results obtained, DoC1 and HC data, using the ten-fold cross-validation approach. To account for the imbalanced target-to-non-target ratio, we use the Area under Curve (AUC) measure to assess the performance. Fig. 1 shows the results for the AUC for each subject in the DoC1 and HC group, as well as the average for each group. As is expected, the results for the patient group (mean 0.64 $\pm$0.11) are significantly lower than for the control group (mean 0.91$\pm$0.07). Also, they exhibit a larger variance (t-test p$\ll$ 0.001, t=5.87). Patients 1-4,6-8,10,11,14,16,17 (12 out of 17) AUCs are well above chance, while Patients 5,9,12 and 15 AUC is not significantly above chance at the 5% level, but close (P5 p=0.061; P9 p=0.058; P12 p=0.056; P15 p=0.057). Only the result of patient 13 is chance (p=0.45). These results are in line with other state-of-the-art approaches studying single-trial classification of DoC patient EEG (see Sec. I for details).

A. Stimulation Block Comparison

The DoC2 study provides a closer insight into the dependency of classification performance and stimulation time. The results, again expressed as the average AUC over ten cross-validation foldings, are presented in Fig. 2. They are split into the individual results of the first and the second stimulation block, and the merged data set. For the latter, all epochs were randomly shuffled, such that each folding subset contains an arbitrary number of epochs from blocks one and two. The average over all patients does hardly differ

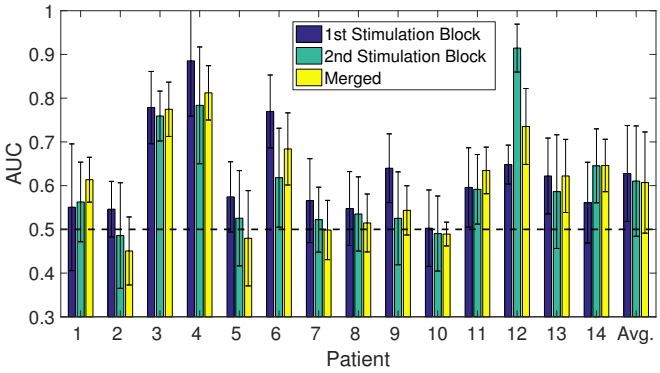


Fig. 2: Results of single-trial classification of data of 14 DoC patients that were stimulated in two blocks.

between the merged and block-wise approaches (merged blocks mean 0.61$\pm$0.12; Block 1 mean 0.63$\pm$0.11; Block 2 mean 0.61$\pm$0.13). Confirming the results from the DoC1 study, the variance between patients is very high. These averaged results, however, mask the interesting patient-wise details. For patients 2,5,7 and 8 the AUC on the merged set is lower than that of the individual blocks. This is in so far remarkable, as more training data, here even double as many epochs, should increase the performance. Even more, while patients 2 (p=0.04) and 5 (p=0.02) achieve an AUC above chance when cross-validating only the first block, the results drop to chance level in the second block (P2 p=0.72, P5 p=0.47). For patient 12, there is an extreme reversed effect. The AUC in the second block is 0.27 larger than in the first block. Patients 1 and 14 also achieve a better performance in the second block with a less pronounced difference. On all other patient sets, the AUC is better in block one than in block two. This could be expected when assuming that the recording and stimulation session is rather exhausting for the patients. However, this does not seem to be necessarily the case. These results clearly motivate to further investigate individual differences related to a patient's mental state and their influence on the classification performance.

B. Frequency Band Variations and Classifier Performance

We use the RCE value introduced in Sec. III-E to assess the influence of variations in specific frequency bands onto the classification performance. We run this process for all three study sets DoC1, DoC2 and HC. RCE is assessed separately for target and non-target epochs. For a first overall assessment, we calculated for each study set the percentage of all RCEs <1. The number of RCE values is 160, 5 frequency bands and 32 channels. Table II lists these results. The results aggregate the contributions of each individual data set in a study, which at this point are not examined individually. There are relevant differences between the three study sets. In the DoC2 set, independent of evaluating the merged or block-wise approach, there is a much higher ratio of misclassifying target epochs. More than 90% of all 160 RCE values are < 1, which corresponds to a very high likelihood of misclassifying a target epoch in a high

TABLE II: Percentages of RCE values <1, i.e., corresponding to a high likelihood of misclassifying epochs in high frequency band variation windows (values in percent).

	DoC1	DoC2	Block1	Block2	HC
Targets	65.0	91.2	94.3	90.6	35.0
Non-Targets	83.1	41.2	52.5	53.1	87.5

variation window. For non-target epochs, this is much lower, with roughly only 50% of RCE values < 1, indicating a lower likelihood. For the other DoC patient set, and even more pronounced for the HC group, this effect is inverted. There, the percentage of RCE values for non-target epochs is much higher, and for target epochs lower. Specifically, in the HC group only 35% of RCE values for target epochs are below 1 and thus, in this group, there is no real correspondence between high frequency band variations and misclassifications of target epochs. However, in the control group, there is a high likelihood of misclassifying non-target epochs. In summary, a first broad assessment of frequency band variations reveals that this measure is most sensitive to the *class* of an epoch, and that this differs between data sets.

For further analysis, we looked into the relation of frequency band variations and misclassifications specifically in the DoC2 study, for the merged set and the two block sets. We focus here on the *alpha* band and look into differences between merged, block one and block two. Fig. 3 shows the detailed results for these sets, plotting the RCEs for the alpha band, individually for each of the 32 channels. We find the previous results confirmed. Again, for target epochs RCE is < 1 for almost all channels and the three sets. For non-target epochs, in contrast, the RCE close to one indicates no clear relation between correct or incorrect classifications with alpha band variations. The smaller RCE for the block two only data, compared to block one data, hints at a possible connection of windows with higher frequency variations and the poorer classification performance for several patients. Although this effect is not consistent for all patients, it indicates a high degree of individual differences, as to be expected with DoC patients.

V. CONCLUSION

We presented an approach to relate the performance of a single-trial classifier with relative frequency bands. We found data set depended effects, where epochs in time windows with high variations in the relative frequency bands tend to be significantly more often misclassified. Interestingly, in some data sets, especially in the DoC2 data, this effect mainly occurs for the target class epochs, while in other data sets, this relationship is reversed. In the future, we will extend our approach to using the introduced RCE value to guide the selection of training samples. We expect that by taking additional measures of a patient's mental state into account for single-trial classification, we can improve its performance to enable BCI communication with DoC patiens.

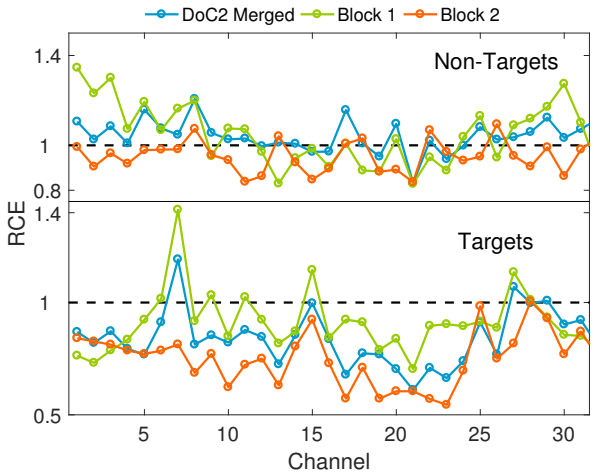


Fig. 3: The RCE value corresponding to the likelihood of misclassifying epochs in phases of high frequency band variations, depicted for the alpha band and all 32 channels.

REFERENCES

[1] J. Sitt, B. Rohaut, A. Gramfort, and L. Naccache, "Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state," *BRAIN*, vol. 137, pp. 2258–2270, 2014.

[2] D. R. Sebastiano, F. Panzica, E. Visani, F. Rotondi, V. Scaioli, M. Leonardi, D. Sattin, L. D'Incerti, E. Parati, L. F. Strambi, and S. Franceschetti, "Significance of multiple neurophysiological measures in patients with chronic disorders of consciousness," *Clinical Neurophysiology*, vol. 126, no. 3, pp. 558 – 564, 2015.

[3] I. Steppacher, S. Eickhoff, J. Todor, M. Kaps, W. Witzke, and J. Kissler, "N400 predicts recovery from disorders of consciousness," *Annals of Neurology*, vol. 73, pp. 594–602, 2012.

[4] D. Cruse, S. Beukema, S. Chennu, J. G. Malins, A. M. Owen, and K. McRae, "The reliability of the n400 in single subjects: Implications for patients with disorders of consciousness," *NeuroImage: Clinical*, vol. 4, pp. 788 – 799, 2014.

[5] L. Farwell and E. Donchin, "Talking off the top of your head: towards a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. S2, pp. 510–523, 1988.

[6] C. Pokorny, D. S. Klobassa, G. Pichler, H. Erlbeck, R. G. Real, A. Kbler, D. Lesenfants, D. Habbal, Q. Noirhomme, M. Risetti, D. Mattia, and G. R. Mller-Putz, "The auditory p300-based single–switch brain–computer interface: Paradigm transition from healthy subjects to minimally conscious patients," *Artificial Intelligence in Medicine*, vol. 59, no. 2, pp. 81–90, 2013.

[7] R. Spataro, A. Heilinger, W. Cho, S. Marchese, D. De Cicco, C. Gregoretti, P. Volanti, C. Guger, and V. La Bella, "Probing consciousness in unresponsive wakefulness syndrome patients with a vibrotactile p300-based brain-computer interface," *Neurology*, vol. 88, no. 16, 2017.

[8] S. Bagnato, C. Boccagni, A. Sant'Angelo, C. Prestandrea, R. Mazzilli, and G. Galardi, "Eeg predictors of outcome in patients with disorders of consciousness admitted for intensive rehabilitation," *Clinical Neurophysiology*, vol. 126, no. 5, pp. 959 – 966, 2015.

[9] A. Finke, K. Essig, G. Marchioro, and H. Ritter, "Toward FRP-based brain-machine interfaces: Single-trial classification of fixation-related potentials," *PLoS ONE*, vol. 11, no. 1, pp. 1–19, 01 2016.

[10] A. Finke, A. Lenhardt, and H. Ritter, "The MindGame: A P300-based brain-computer interface game," *Neural Networks*, vol. 22, pp. 1329–1333, 2009.

[11] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, February 2004.

[12] J. Schaefer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, p. Article 32, 2005.