Evaluation of electroencephalography analysis methods*

Dominik Wetzel[1†], Nico Spahn[1], Martin Heilemann[1], Marcus M. Löffler[1],
Markus Seidel[1], Silke Kolbig[1], Dirk Winkler[2]

***Abstract*— The extraction of expressive features from an electroencephalography (EEG) signal is necessary for classification of movement and movement imagination of the limbs. We introduce different preprocessing and feature extraction algorithms for this purpose and develop an algorithm that selects features by their feature importance. This selection is used as an evaluation measure for features, their preprocessing algorithms and the EEG electrodes. Our results show that most influential features for signal interpretation are: common spatial patterns, fractal dimensions, as well as, variance and standard deviation of the preprocessed data. We show that preprocessing with continuous wavelet transforms outperforms the other tested preprocessing algorithms. Furthermore, we show that high gamma frequencies (70-90 Hz) contain more information than the lower μ-rhythms (8-12 Hz) where event-related-desynchronization (ERD) is known to occur. The important EEG electrodes for this classification task are located in the left and right back of the motor-cortex. The proposed algorithm can be further used to create subject-specific and performance models for real-time classification.**

I. Introduction

Real-time brain-computer interface (BCI) applications need fast and stable methods for signal preprocessing and feature extraction. Electroencephalography signals are highly subject-specific and show a poor signal-to-noise ratio due to their origin from beneath multiple biological layers. For classification or regression tasks, it is crucial to calculate expressive features of the signal in order to extract the important information. For example ERD occurs during movement or the imagination of movement of the upper or lower limbs [1]–[3]. This seems to be a meaningful feature of detecting movement imagination in across-trial averaging [1], but therefore it has limited potential for real-time applications. Many approaches show alternative feature extraction methods [4], [5] which can be considered.

In this work several preprocessing and feature extraction algorithms were selected that could be utilized for real-time analysis. Therefore, a tool was developed to rate the features for each individual subject. and select the best analysis algorithms.

II. Experimental setup

Experiments were performed with 13 healthy adult subjects, where written informed consent to participate in this study was obtained. Seven male (mean age: 30.9 ± 3.0) and six female subjects (mean age: 27.2 ± 5.6) were analyzed. Each subject participated in two sessions. The time between the sessions was at least 3.5 hours (for one subject the time between sessions was about 3 weeks). In each session, two experiments were performed: the movement experiment (ME) where the subjects were asked to move the limb shown on a standard 60 Hz display (left arm, right arm, left leg, right leg) and the imagine experiment (IE) where the subjects should just imagine the movement. The duration of an experiment was approx. 10 minutes. The experimental flowchart is shown in Fig. 1. These experiments were conducted in accordance to the principles of the Declaration of Helsinki (2000) [6]. The study was purely observational.

After an initial five seconds, a reference image (see Fig. 2a) was shown. Two to five seconds later a class image (one of Figs. 2b to 2e) with an acoustic signal was presented and an electrical signal (trigger) was sent to the measuring devices. The subject was instructed to move (or imagine the movement of) the shown limb. The image vanished after two seconds. The screen was left blank for five to ten seconds and the next trial began by showing the reference image. This procedure was repeated 40 times so that every class was presented ten times. The sequence of the class images was randomized. The above-mentioned trigger was used to synchronize the devices and to find the beginning of the trials in the data.

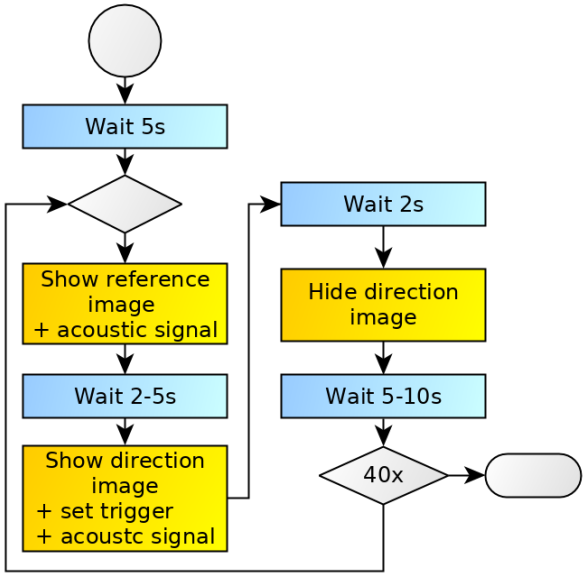


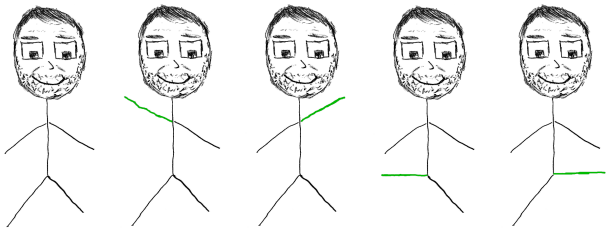
Fig. 1: The flowchart for the ME and IE experiments

*This research paper was funded by the European Social Fund (ESF) with grant number 100270152.

[1]Affiliated with the Faculty of Physical Engineering/Computer Sciences, University of Applied Sciences Zwickau, 08012 Zwickau, Germany

[2]Affiliated with the Department of Neurosurgery, Faculty of Medicine, Leipzig University, 04103 Leipzig, Germany

[†]Email: dominik.wetzel@fh-zwickau.de



(a) (b) (c) (d) (e)

Fig. 2: (a) reference image, (b)-(e) class images (left arm, right arm, left leg, right leg)

Before the beginning of the measurements the participants tried the movements and saw the reference and some class images.

During the experiments, the EEG-signal was measured with g.SCARABEO active Ag/AgCl-electrodes with the g.HIamp amplifier from Guger Technologies (g.tec) with 256 Hz sampling rate. The signal was preprocessed with a 48-52 Hz Butterworth notch filter, as well as, a 0.5-100 Hz 8th order Butterworth bandpass filter. We used electrodes positioned as shown in Fig. 3 of the 10–10 system. For each electrode, the impedance was measured before a session and the value was always below 50 $k\Omega$. Furthermore, during MEs, we simultaneously recorded the movement of the limbs using acceleration sensors to determine the correct start and end times for the movement. After each session we cleaned and disinfected the devices.

To evaluate the IE we supposed similar inherent timing for each subject where we approximated the start of movement imagination and the time for movement imagination by taking the mean of the previous ME data for the real movement for each class. In the end, a labeled data set for the real movement and the imagination of the movement was acquired. Fig. 4 shows a schematic of the experimental setup.

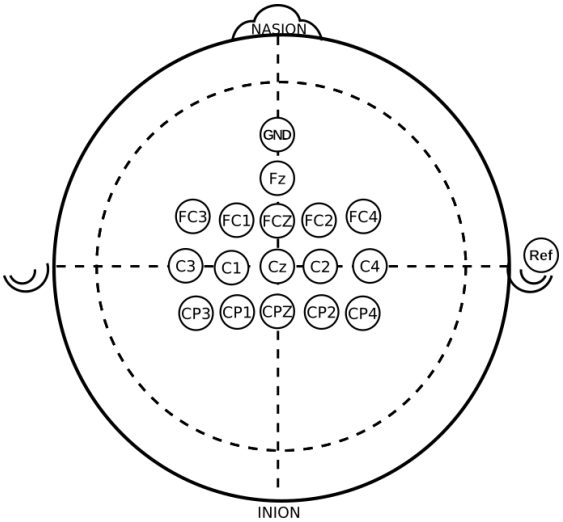


Fig. 3: Electrode positions of the 10–10 system

III. Methods

Here we introduce a methodology to evaluate features for real-time EEG analysis. Therefore a series of many different features and their ranking is calculated. This provides the opportunity to select the best ranked features and to use them for online evaluation.

We use a sliding window technique with window size $w = 85$ samples (≈ 332 ms)[‡] and an overlap of 70 samples. We calculate features for each window in two steps. First, the signal is preprocessed with different algorithms to reduce the noise of the signal and secondly, features are calculated for each preprocessed data stream (see Fig. 5).

A. Preprocessing algorithms

Besides keeping the original raw signal unchanged as one variant, we use several variants of two different types of algorithms for preprocessing to get numerous filtered copies of each signal.

a) Bandpower: is the power within a frequency band. To calculate it, we apply a bandpass filter and square the signal. We use three different distinct frequency band-ranges where each range is split into smaller frequency bands as shown in table I. The 8-12 Hz frequency range is known for containing the ERDs [11]. The band-range about ≈ 40 Hz is also stated in literature for containing movement information [12], [13]. We also consider the high gamma frequency ranges (70-90 Hz) as they seem to play an important role for movement classification [13]–[15].

b) Continuous Wavelet Transform (CWT): It is well known that wavelet transforms have good properties in the domain where non-periodicity and non-stationarity have to be assumed, which is true for EEG-signals [14]. In our research,

[‡]These values are motivated by the fact that approximately 300 ms after a decision or stimulus, a unique pattern can be seen in the EEG (P300) [7], [8], which can be used for controlling BCI applications [9], [10].

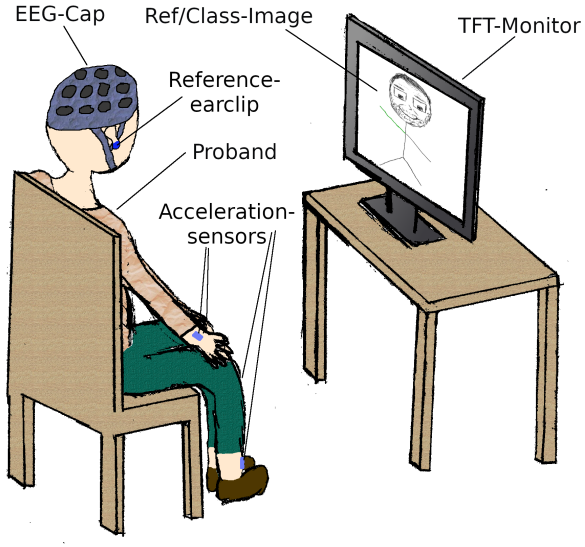


Fig. 4: Schematic figure for the experimental setup

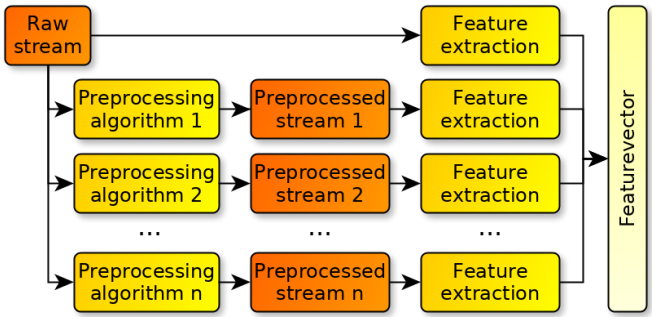


Fig. 5: The workflow for feature calculation. This is done for each window of each electrode signal.

we use a mexican-hat wavelet implemented in PyWavelets [16] which is defined in the time domain as:

$$\Psi(t) = \frac{2}{\sqrt{3}\sqrt[4]{\pi}}(1-t^2)e^{-\frac{t^2}{2}} \tag{1}$$

Due to the shape of the mexican-hat wavelet, it is possible to approximate a frequency with a given scale of the wavelet (see Fig. 6). In our research we use the scales shown in table II (which approximately resembles the frequencies of the bandpower).

B. Feature extraction

Feature extraction algorithms can be separated into mainly two classes: Single electrode features (that are calculated for each electrode) and multielectrode features (that use combinations of electrodes for a single feature).

1) Single electrode features: We calculated 11 different features on each data stream. These are nine statistical features, fractal dimension (FD) and running variation coefficient (RVC).

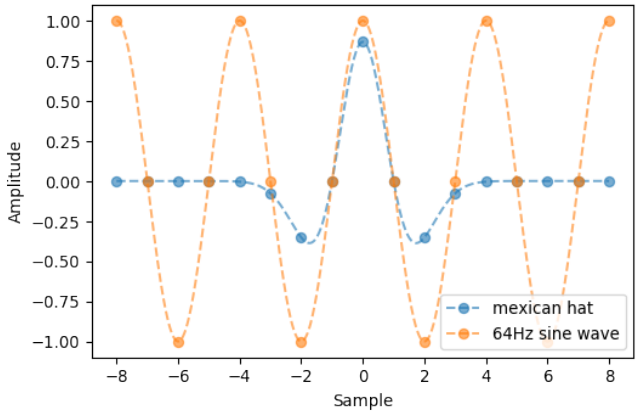


Fig. 6: The mexican-hat wavelet with scale 1 and a sine wave with frequency 64 Hz, both discrete sampled with 256 Hz. It can be seen that the peak of the wavelet approximates the sine wave at this position.

a) Statistical features: The nine statistical measures are:

- Minimum: $min(x)$
- Maximum: $max(x)$
- Range: $max(x) - min(x)$
- Arithmetic mean: $\bar{x} = \frac{1}{n}\sum_{i-n}^{i} x_i$
- Variance: $\sigma^2 = \frac{1}{n}\sum_{i-n}^{i}(x_i - \bar{x})^2$
- Standard deviation: σ
- 25-percentile
- 50-percentile (median)
- 75-percentile

where x_i is the individual data point and x is the windowed data set.

b) Fractal dimension: is a measure of complexity of the signal [17]. We use Katz' method [17] with slight adjustments as it is very fast to calculate compared with other algorithms [18]. The formula used for the FD is therefore given as

$$FD = \frac{ln(L)}{ln(d)} \tag{2}$$

$$L = \sum_{i}^{w}\sqrt{(P_x^i - P_x^{i+1})^2 + (P_y^i - P_y^{i+1})^2} \tag{3}$$

$$d = \max_{i=1\ldots w}\sqrt{(P_x^1 - P_x^i)^2 + (P_y^1 - P_y^i)^2} \tag{4}$$

where L is the length of the signal curve, d is the farthest distance between the start point (P^1) and any of its predecessors and w is the window length. P_x^i refers to the x-component of the i-th point, while P_y^i refers to the y-component respectively (in the respective unit of the preprocessing algorithm). The distance in x-direction was set to 1 for each consecutive point, which simplifies the formula for L and d to:

$$L = \sum_{i}^{w}\sqrt{1 + (P_y^i - P_y^{i+1})^2} \tag{5}$$

$$d = \max_{i=1\ldots w}\sqrt{(i-1)^2 + (P_y^1 - P_y^i)^2} \tag{6}$$

c) Running variation coefficient: We calculate the variation coefficient for the window, but calculate the mean ($\bar{x}$) not only for the current window but also for the n windows before. This results in the following formula for the RVC:

$$RVC(x) = \frac{\sigma}{\bar{x}(n)}, \text{ with } \bar{x}(n) = \frac{1}{n}\sum_{i=-n}^{0}\bar{x}_i \tag{7}$$

2) Multielectrode features: We consider only one multielectrode feature, namely the common spatial pattern (CSP), which was first introduced for EEG data in [19] and first used for BCI systems for binary classification problems in [20]. Further it was expanded for multiclass problems in [21]. We use a Python implementation from the MNE library [22], [23] to calculate the feature. The CSP is an algorithm that approximately maximizes mutual information of independent components of the signals and the class labels [21]. For our research, we transform the data into CSP-space and extract the average power of the eight best CSP features.

C. Roundup amount of features

Considering the preprocessing algorithms, the number of electrodes and feature-classes, we can calculate the amount

TABLE I: The used frequency bands for bandpower calculation. First column shows abbreviation and main band-range. Values are in Hz

BP1	8–12.5	8–9	8.5–9.5	9–10	9.5–10.5	10–11	10.5–11.5	11–12	11.5–12.5
BP2	39–47	39–40	40–41	41–42	42–43	43–44	44–45	45–46	46–47
BP3	70–95	70–80	75–85	80–90	85–95				

TABLE II: The used scales for the CWT. The approximated frequency in Hz is shown in braces.

CWT1	5–9	5.00	(13)	6.00	(11)	7.00	(9)	8.00	(8)	9.00	(7)
CWT2	1.35–2.15	1.35	(47)	1.55	(41)	1.75	(37)	1.95	(33)	2.15	(30)
CWT3	0.7–0.9	0.70	(91)	0.75	(85)	0.80	(80)	0.85	(75)	0.90	(71)

of features used. For the preprocessing we have 20 bandpowers, 15 CWTs and one raw stream, resulting in 36 data streams for each of the 16 raw signals coming from the 16 electrodes. For each of the streams there are nine statistical features, one FD and one RVC. CSP adds eight features per data stream. This sums up to $11{\cdot}36{\cdot}16{+}8{\cdot}36 = 6624$ features per window. In order to reduce the amount of data and to discriminate features that contain information from features that do not, we use methods from the domain of machine learning. These are described in the following section.

D. Ranking and evaluation

To evaluate our proposed methods, we use the Random Forests classification algorithm [24] implemented in scikit-learn [25]. This algorithm provides an instrument to sort a large number of features by their importance and hence to single out the most important ones. The training of a forest with our labeled data provides a feature importance vector, which can be used as a ranking function for the features. Another important property of Random Forests is that they are feature-scale invariant (because of the underlying usage of decision trees) so the features do not need to be scaled into a specified range.

For evaluation, we train the Random Forest and sort the features by their feature importance (resulting in the sorted feature vector r with the sorted feature importance vector f which sums to 1). Afterwards we define a cutoff feature percentage (C_p) and define our feature measure m_p as follows:

$$m_p(r_i) \quad = \quad \begin{cases} f_i: & \sum_{j=1}^{i} f_j \leq \frac{C_p}{100} \\ 0: & \text{otherwise} \end{cases} \tag{8}$$

For each feature we additionally store the data stream and the electrode where it originates from. This allows us to determine which features, which preprocessing algorithms and which electrodes are important. For our research we set $C_p = 10\%$ to limit the amount of selected features.

IV. Results and Discussion

To evaluate the importance of the features, preprocessing methods and electrodes, we calculate the feature measure in two ways. Firstly, for all features and secondly, for the subset of all features except CSP because CSP uses all electrodes which makes the electrode distribution not as expressive.

Using all features: Figs. 7a and 7d shows the importance of the features. It can be seen that the most important feature is CSP followed by FD. In ME, variance, standard deviation and range play a role, whereas in IE range is less important. RVC is not considered once and has therefore no meaning for the classification in this setting.

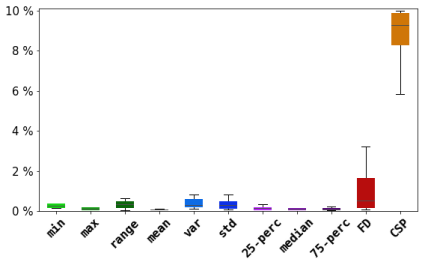
For the preprocessing, we can see in Figs. 7b and 7e that the CWT preprocessing is more important than the other algorithms. Furthermore, the CWT with higher frequencies (39-90 Hz) contains more useful information. This is true for both ME and IE, although the CWT around 8-13 Hz adds approximately equally to the classification in IE. Such findings maybe due to the fact that electromyographic (EMG) signals corrode the pure brain signal during thinking and movement [26], [27]. For our use-case, it does not matter much if the measured data is from EMG (that is emitted by thinking) or the real brain data as the subjects should also be able to control these signals for their intentions. Unexpectedly here, the bandpower preprocessing in the lower frequency bands plays a minor role [1]. Also the raw data stream does contain useful features.

Figs. 7c and 7f show the mean electrode distribution. As can be seen, every electrode contributes approximately equally to the classification, which is not surprising as CSP uses all electrodes. Therefore we looked into the distribution without the calculation of CSP.

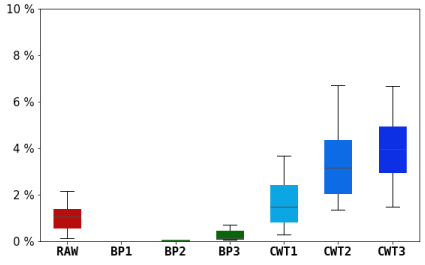
Without CSP: In this setting the median amount of features used to reach the 10% hurdle given by the cutoff parameter C_p is about nine times higher than in the setting with using all features (see Fig. 9) which in turn means that CSP contributes significantly to the classification.

On the feature side, Figs. 8a and 8d show that FD contributes the most followed by the minimum, variance and standard deviation in ME. The contributions are similar in IE but also include the maximum . What is evident is that the mean and RVC do not add to the classification and thus are useless for this task.

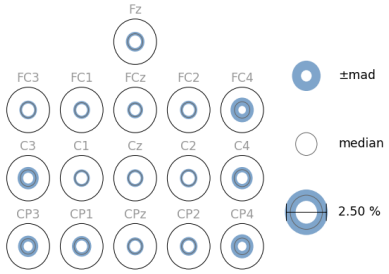
The distribution of the preprocessing algorithms stays the same (see Figs. 8b and 8e), only CWT in the range about 70-90 Hz is much more important and the bandpower in the lower (8-13 Hz) and the upper range (70-90 Hz) contributed. Interestingly, bandpower in the range of 39-47 Hz adds nothing to the classification. As can be seen in Figs. 8c and 8f, the outer electrodes contain more useful features for this classification task than the electrodes inside. This is not



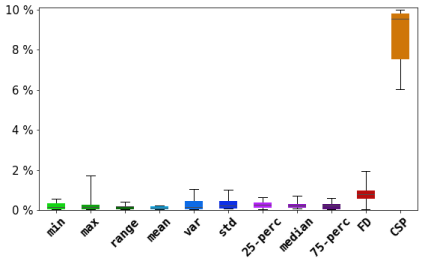
(a) Used features in ME



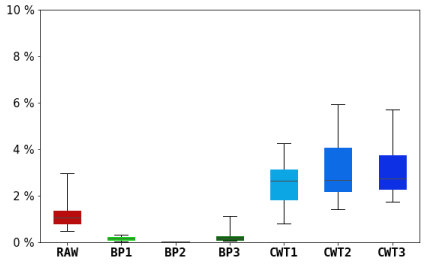
(b) Used preprocessing in ME



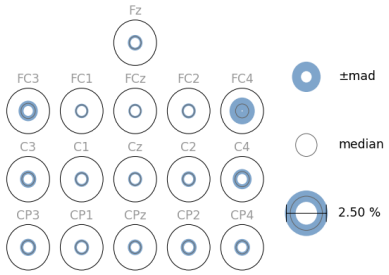
(c) Used electrodes in ME



(d) Used features in IE

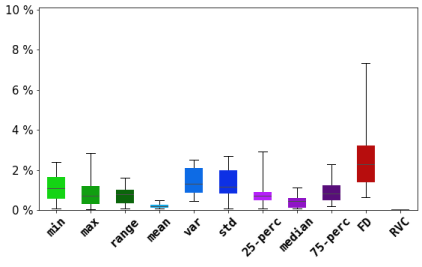


(e) Used preprocessing in IE

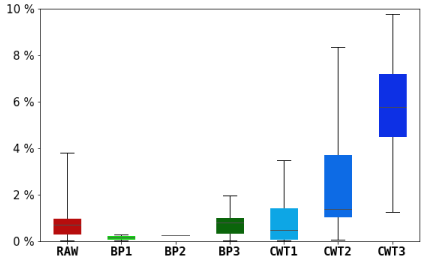


(f) Used electrodes in IE

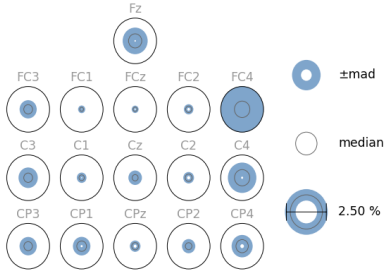
Fig. 7: Evaluation of all features. The whiskers of the boxplots are the range of the feature importance of the given view. (a), (d): View on the importance by features. (b), (e): View on the importance by preprocessing types. (c), (f): View on the importance by the used electrodes. Diameter of the gray circles is the median importance of this electrode and the blue ring is its median absolute deviation (mad) of the importance.



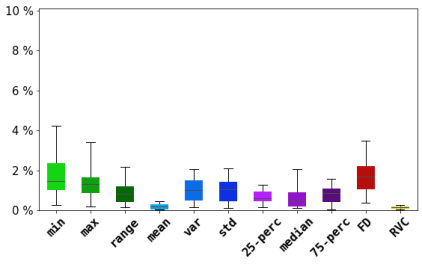
(a) Used features in ME



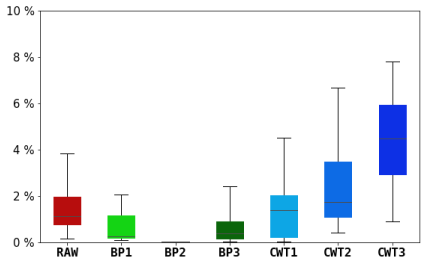
(b) Used preprocessing in ME



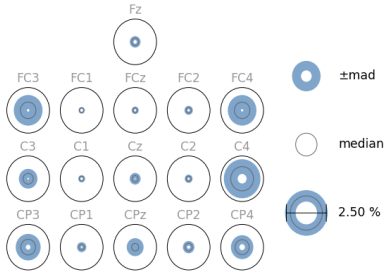
(c) Used electrodes in ME



(d) Used features in IE



(e) Used preprocessing in IE



(f) Used electrodes in IE

Fig. 8: Evaluation of features without CSP. The whiskers of the boxplots are the range of the feature importance of the given view. (a), (d): View on the importance by features. (b), (e): View on the importance by preprocessing types. (c), (f): View on the importance by the used electrodes. Diameter of the gray circles is the median importance of this electrode and the blue ring is its median absolute deviation (mad) of the importance.

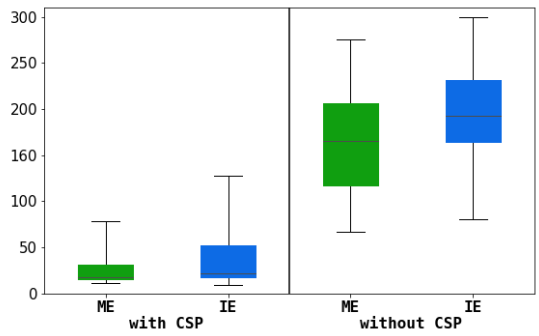


Fig. 9: The amount of features used to reach the 10% hurdle. Left shows with all features and right without CSP features.

surprising as it is known that the left and right hemisphere correspond to the right and left side, respectively, of the body which fits to our analysis. The surprising finding is that even the electrode in the front (Fz) is important in some cases. Also the back of the head (CP3 to CP4) contributes meaningfully to our task.

Real movement and imagination: The difference between ME and IE is not as significant as expected. The amount of features increases just slightly from ME to IE as can be seen in Fig. 9. Even the distribution of the features and the electrodes differ only slightly. The distribution of the preprocessing algorithms also did not change much except for the usage of CWT1 as its weight is higher in IE than in ME in most cases. This may be due to less EMG artifacts in the data and therefore a higher relevance of the ERD which should be present in the frequency range of CWT1.

V. Conclusion

We discussed two well-known preprocessing methods as well as some feature extraction methods. An algorithm was implemented that allows the determination of subject-specific preprocessing methods, features and best performing electrode positions. This algorithm can be utilized to create powerful and performance feature models for each individual subject which reduces preparation and computation time for the subjects during consecutive sessions. We use the feature importance measure of Random Forest classifiers for dimensional reduction as these are feature scale invariant as opposed to other algorithms used in literature (e. g. [4], [28]).

Our findings indicate that preprocessing the signals with CWT is superior to preprocessing them with bandpower. Another important finding is that most information for our classification task seems to be inside the high gamma band (70-90 Hz), followed by the 39-45 Hz range. Further, we can say that CSP has the most impact on classification. FD also seems to be a very meaningful feature. Other useful features are variance, standard deviation and minimum, but it is necessary to use features that work best for the individual subject so other features may still be considered, with the exception on RVC and mean as they add nothing to the classification. Important electrodes for this classification task are Fz, FC3, FC4, C3, C4, CP3 and CP4 which corresponds to the left and right hemisphere. For further research we may additionally consider CP1, CPz and CP2 as they seem to play some role.

In our next research steps, we will use these results and perform real-time classification of the described five-class experiment. Therefore, we not only want to rely on the algorithms provided, but also on brain plasticity using neurofeedback to achieve good results.

References

[1] G. Pfurtscheller and F. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles", *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842–1857, Nov. 1999, ISSN: 13882457. DOI: 10.1016/S1388-2457(99)00141-8.

[2] A. A. Aleksandrov and S. M. Tugin, "Changes in the Mu Rhythm in Different Types of Motor Activity and on Observation of Movements", *Neuroscience and Behavioral Physiology*, vol. 42, no. 3, pp. 302–307, Mar. 2012, ISSN: 0097-0549. DOI: 10.1007/s11055-012-9566-2.

[3] C. Llanos, M. Rodriguez, C. Rodriguez-Sabate, I. Morales, and M. Sabate, "Mu-rhythm changes during the planning of motor and motor imagery actions", *Neuropsychologia*, vol. 51, no. 6, pp. 1019–1026, May 2013, ISSN: 0028-3932. DOI: 10.1016/J.NEUROPSYCHOLOGIA.2013.02.008.

[4] L. Vega-Escobar, A. Castro-Ospina, and L. Duque-Munoz, "Feature extraction schemes for BCI systems", in *2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA)*, IEEE, Sep. 2015, pp. 1–6, ISBN: 978-1-4673-9461-1. DOI: 10.1109/STSIVA.2015.7330455.

[5] A. S. Al-Fahoum and A. A. Al-Fraihat, "Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains", *ISRN Neuroscience*, vol. 2014, pp. 1–7, 2014, ISSN: 2314-4661. DOI: 10.1155/2014/730218.

[6] "Declaration of Helsinki : ethical principles for medical research involving human subjects : current topics", *WHO drug information 2000 ; 14(3) : 160-162*, 2000.

[7] E. Donchin, "Surprise!? Surprise?", *Psychophysiology*, vol. 18, no. 5, pp. 493–513, Sep. 1981, ISSN: 0048-5772. DOI: 10.1111/j.1469-8986.1981.tb01815.x.

[8] S. Hillyard and M. Kutas, "Electrophysiology of Cognitive Processing", *Annual Review of Psychology*, vol. 34, no. 1, pp. 33–61, Jan. 1983, ISSN: 0066-4308. DOI: 10.1146/annurev.ps.34.020183.000341.

[9] L. 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, no. 6, pp. 510–523, Dec. 1988, ISSN: 0013-4694. DOI: 10.1016/0013-4694(88)90149-6.

[10] A. Pinegger, H. Hiebel, S. C. Wriessnegger, and G. R. Müller-Putz, "Composing only by thought: Novel application of the P300 brain-computer interface", *PLOS ONE*, vol. 12, no. 9, P. Sardo, Ed., e0181584, Sep. 2017, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0181584.

[11] G. Pfurtscheller, C. Neuper, and G. Krausz, "Functional dissociation of lower and upper frequency mu rhythms in relation to voluntary limb movement", *Clinical Neurophysiology*, vol. 111, no. 10, pp. 1873–1879, Oct. 2000, ISSN: 1388-2457. DOI: 10.1016/S1388-2457(00)00428-4.

[12] G. Pfurtscheller, D. Flotzinger, and C. Neuper, "Differentiation between finger, toe and tongue movement in man based on 40 Hz EEG", *Electroencephalography and Clinical Neurophysiology*, vol. 90, no. 6, pp. 456–460, Jun. 1994, ISSN: 0013-4694. DOI: 10.1016/0013-4694(94)90137-6.

[13] T. Ball, E. Demandt, I. Mutschler, E. Neitzel, C. Mehring, K. Vogt, A. Aertsen, and A. Schulze-Bonhage, "Movement related activity in the high gamma range of the human EEG", *NeuroImage*, vol. 41, no. 2, pp. 302–310, Jun. 2008, ISSN: 1053-8119. DOI: 10.1016/J.NEUROIMAGE.2008.02.032.

[14] W. Klonowski, "Everything you wanted to ask about EEG but were afraid to get the right answer.", *Nonlinear biomedical physics*, vol. 3, no. 1, p. 2, May 2009, ISSN: 1753-4631. DOI: 10.1186/1753-4631-3-2.

[15] F. Darvas, R. Scherer, J. Ojemann, R. Rao, K. Miller, and L. Sorensen, "High gamma mapping using EEG", *NeuroImage*, vol. 49, no. 1, pp. 930–938, Jan. 2010, ISSN: 1053-8119. DOI: 10.1016/J.NEUROIMAGE.2009.08.041.

[16] G. Lee, R. Gommers, F. Wasilewski, K. Wohlfahrt, A. O'Leary, and H. Nahrstaedt, *PyWavelets - Wavelet Transforms in Python*, 2006. [Online]. Available: https://github.com/PyWavelets/pywt (visited on 10/18/2018).

[17] M. J. Katz, "Fractals and the analysis of waveforms", *Computers in Biology and Medicine*, vol. 18, no. 3, pp. 145–156, Jan. 1988, ISSN: 0010-4825. DOI: 10.1016/0010-4825(88)90041-8.

[18] U. Güçlü, Y. Güçlütürk, and C. K. Loo, "Evaluation of fractal dimension estimation methods for feature extraction in motor imagery based brain computer interface", *Procedia Computer Science*, vol. 3, pp. 589–594, Jan. 2011, ISSN: 1877-0509. DOI: 10.1016/J.PROCS.2010.12.098.

[19] Z. Koles, "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG", *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 6, pp. 440–447, Dec. 1991, ISSN: 0013-4694. DOI: 10.1016/0013-4694(91)90163-X.

[20] H. Ramoser, J. Muller, and G. Pfurtscheller, "Optimal spatial ltering of single trial EEG during imagined hand movement", *Rehabilitation*, vol. XX, pp. 100–120, 1998.

[21] M. Grosse-Wentrup and M. Buss, "Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction", *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1991–2000, Aug. 2008, ISSN: 0018-9294. DOI: 10.1109/TBME.2008.921154.

[22] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. Hämäläinen, "MEG and EEG data analysis with MNE-Python", *Frontiers in Neuroscience*, vol. 7, p. 267, Dec. 2013, ISSN: 1662453X. DOI: 10.3389/fnins.2013.00267.

[23] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen, "MNE software for processing MEG and EEG data", *NeuroImage*, vol. 86, pp. 446–460, Feb. 2014, ISSN: 1053-8119. DOI: 10.1016/J.NEUROIMAGE.2013.10.027.

[24] L. Breiman, "Random Forests", *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, ISSN: 08856125. DOI: 10.1023/A:1010933404324.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825–2830, 2011, ISSN: 1533-7928.

[26] E. M. Whitham, K. J. Pope, S. P. Fitzgibbon, T. Lewis, C. R. Clark, S. Loveless, M. Broberg, A. Wallace, D. DeLosAngeles, P. Lillie, A. Hardy, R. Fronsko, A. Pulbrook, and J. O. Willoughby, "Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG", *Clinical Neurophysiology*, vol. 118, no. 8, pp. 1877–1888, Aug. 2007, ISSN: 1388-2457. DOI: 10.1016/J.CLINPH.2007.04.027.

[27] E. M. Whitham, T. Lewis, K. J. Pope, S. P. Fitzgibbon, C. R. Clark, S. Loveless, D. DeLosAngeles, A. K. Wallace, M. Broberg, and J. O. Willoughby, "Thinking activates EMG in scalp electrical recordings", *Clinical Neurophysiology*, vol. 119, no. 5, pp. 1166–1175, May 2008, ISSN: 13882457. DOI: 10.1016/j.clinph.2008.01.024.

[28] P. J. García-Laencina, G. Rodríguez-Bermudez, and J. Roca-Dorda, "Exploring dimensionality reduction of EEG features in motor imagery task classification", *Expert Systems with Applications*, vol. 41, no. 11, pp. 5285–5295, Sep. 2014, ISSN: 0957-4174. DOI: 10.1016/J.ESWA.2014.02.043.