Hasenhütl Michael

Decoding Combined EEG Trials of Several Individuals and Electrode Systems

Bachelor's Thesis

to achieve the university degree of

Bachelor of Science

Bachelor's degree programme: Biomedical Engineering

submitted to

Graz University of Technology

Supervisor

Univ.-Prof. Dipl.-Ing. Dr.techn. Gernot Müller-Putz

Institute of Neural Engineering
Head: Univ.-Prof. Dipl.-Ing. Dr.techn. Gernot Müller-Putz

Graz, October 2022

This document is compiled with pdfLaTeX2e and `Biber`.

The LaTeX template from Karl Voit is based on KOMA script and can be found online: https://github.com/novoid/LaTeX-KOMA-template

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used.

Date

Signature

Zusammenfassung

Das Greifen von Gegenständen ist ein wesentlicher Bestandteil des täglichen Lebens. Nach Verlust dieser Fähigkeit können Brain-Computer-Interfaces (BCI) zur Kompensation beitragen. Neuere Studien konnten Elektroenzephalogramme (EEG) hinsichtlich Bewegungsabsichten dekodieren. Die vorliegende Arbeit konzentriert sich auf die Gemeinsamkeiten im EEG mehrerer Personen bei gleicher Bewegungsintention, sowie auf die Kalibrierung und das Testen von EEG-Klassifikatoren mit den zusammengefassten Daten mehrerer Elektrodensysteme, nämlich dem wasserbasierten System EEG-Versatile^{TM}, dem Trockenelektroden-Headset EEG-Hero^{TM} und dem gel-basierten Goldstandard (g.USBamp/g.Ladybird). Die Ergebnisse zeigten, dass der Unterschied zwischen Ruhe und Bewegungsintentionen in allen Szenarien (bei Untersuchung teilnehmer- und systemübergreifender Daten) unterscheidbar bleibt. Allerdings war die Dekodierbarkeit zwischen den Bewegungsklassen „lateraler Griff" und „Handflächengriff" nicht signifikant. Um zwischen Ruhe und den verschiedenen Bewegungszuständen zu unterscheiden, wurde ein Multiklassen-Klassifikator basierend auf linearer Diskriminanzanalyse (LDA) implementiert. Dieser wurde mit den zusammengefassten EEG-Daten mehrerer Probanden trainiert. Die durchschnittliche Spitzengenauigkeit beim Testen mit Aufzeichnungen von neuen Probanden betrug für das gel-basierte System 50,7% (4,6% STD), 51,3% (5,6% STD) für das wasserbasierte und 48,3% (5% STD) für das Trockenelektrodensystem. Um zu prüfen, wie gut ein kalibrierter Klassifikator auf mehreren Systemen (mit unterschiedlichen Elektrodentypen) eingesetzt werden kann, wurde er zunächst mit den kombinierten Daten aller drei Systeme trainiert. Die durchschnittliche Spitzengenauigkeit lag bei 49,3% (6% STD) beim Testen mit neuen Daten des gel-basierten Systems, bei 50,3% (5,2% STD) für das wasserbasierte System und bei 46,6% (4,6% STD) für das Trockenelektrodensystem. Anschließend wurde anhand der zusammengefassten Daten von nur zwei Systemen trainiert und mit dem verbleibenden System getestet. Hier ergab sich eine durchschnittliche Spitzengenauigkeit von 47,6% (4,9% STD) für das bisher ungesehene, auf Gel basierende Testsystem, 50,7% (5,9% STD) für das Wasserbasis-System und 44,6% (4,8% STD) bei Verwendung des Trockenelektroden-Headsets als Testsystem.

Abstract

Reaching to grasp objects is an essential part of everyday life. To be able to replace this function if taken away, the idea of brain-computer interfaces (BCI) arises. Recent studies based on movement intention detection have shown decodable neural correlates in the electroencephalogram (EEG). This work focuses on the common neural properties of the same movement intentions by different persons, as well as the calibration and testing of EEG classifiers with the combined data of several recording systems, namely the water-based system EEG-Versatile^{TM}, the dry-electrodes headset EEG-Hero^{TM} and a gel-based gold standard (g.USBamp/g.Ladybird). Results showed that the difference between rest and movement intentions stays distinguishable in all scenarios (cross-participant and cross-system data). However, with the implemented setup decodability of the movement classes "lateral grasp" and "palmar grasp" was not significant. In order to distinguish between the different classes of movement conditions and rest a single-trial multiclass linear discriminant analysis (LDA) classifier was implemented. To analyze similarities in the EEG recordings of different persons the classifier was trained with the combined EEG data of several individuals. Grand average peak accuracy calculated on the unseen data of a new participant yielded for the gel-based electrode system 50.7% (4.6% STD), 51.3% (5.6% STD) for the water-based electrode system and 48.3% (5% STD) for the dry electrode system. To inspect how well a calibrated classifier can be used on several systems (with different electrode types) it was first trained with the combined data of all three systems. Grand average peak accuracy was 49.3% (6% STD) when calculated on unseen data from the gel-based system, 50.3% (5.2% STD) for testing on the water-based system and 46.6% (4.6% STD) for the dry-electrode system. Then it was trained using the combined data of just two systems and tested on the remaining system. Here the grand average peak accuracy calculated on the unseen gel-based testing system yielded 47.6% (4.9% STD), 50.7% (5.9% STD) for the water-based testing system and 44.6% (4.8% STD) when using the dry-electrode headset as testing system.

Contents

Zusammenfassung iv

Abstract v

1 Introduction 1
- 1.1 Related Work . . . 1
- 1.2 Problem Description . . . 2

2 Methods 4
- 2.1 Data Preprocessing . . . 4
- 2.2 Single-Trial Multiclass Classification . . . 5
- 2.3 Cross-Participant Classification . . . 6
- 2.4 Cross-System Classification . . . 6
  - 2.4.1 Training and Testing on Data of all Systems . . . 6
  - 2.4.2 Testing on Data of a Previously Unseen System . . . 7
- 2.5 Removing Eye Artifacts . . . 7
- 2.6 The LDA Classifier . . . 8
- 2.7 Covariance Regularization . . . 10
- 2.8 Adjusting the Chance Level . . . 11
- 2.9 Data Normalization . . . 11

3 Results 13
- 3.1 Cross-Participant Classification . . . 13
- 3.2 Cross-System Classification . . . 20
  - 3.2.1 Training All Recording Systems . . . 20
  - 3.2.2 Using a Separate Testing System . . . 24

4 Discussion 29
- 4.1 Confirmations . . . 29
- 4.2 Limitations . . . 30
- 4.3 Remarks . . . 31

5 Conclusion 32

Bibliography 33

List of Figures

1.1 Experimental setup (Schwarz, Escolano, et al., 2020) 3

2.1 Ambiguous regions in multiclass-LDA. (Duda, 2001) 9

3.1 Results of cross-participant classification on gel-based recordings 14
3.2 Results of cross-participant classification on water-based recordings 16
3.3 Results of cross-participant classification on dry-system recordings 18
3.4 Results of all-system classification 21
3.5 Results of cross-system classification when testing on unseen gel-based recordings 25
3.6 Results of cross-system classification when testing on unseen water-based recordings 26
3.7 Results of cross-system classification when testing on unseen dry-system recordings 27

1 Introduction

Grasping objects is one of the most fundamental functions for interacting with the environment. Hence, if taken away, it is a human need to regain this ability. One possible way could be a brain computer interface (BCI) (Wolpaw et al., 2002) which non-invasively uses live electroencephalography (EEG) to guess the grasping intentions of its user. This scenario has been widely studied (Úbeda et al., 2017; López-Larraz et al., 2014; Suwannarat, Pan-Ngum, and Israsena, 2018; Pei et al., 2020). Schwarz, Escolano, et al., 2020 show positive results for several recording systems. This work investigates the transition of a calibrated single-trial classifier between individuals and recording systems.

1.1 Related Work

This work can be seen as a continuation of the study by Schwarz, Escolano, et al., 2020. Their aim was to analyze EEG data recorded during rest and natural reach and grasp actions. The setup for lateral and palmar grasp conditions can be seen in figure 1.1. After splitting the recorded data into trials per condition (rest, lateral grasp and palmar grasp), they trained a separate classification model for each study participant using the respective trials. Applying the model on previously unseen test data of the same participant showed that the conditions are in fact decodable and the results are statistically significant.

The EEG data was recorded using 3 different systems: One gel-based (g.tec USBamp/g.tec Ladybird system, g.tec medical engineering GmbH, Austria), one water-based (EEG-Versatile^{TM}, Bitbrain, Spain) and one dry system (EEG-Hero^{TM}, Bitbrain, Spain). The 45 recordings (15 participants per system) have been published as open access data for further analysis by the scientific community and were used in this work. For comparability the single-trial

classification approach of the related study was reproduced as accurately as possible and then applied on the new problems as described in the following section.

1.2 Problem Description

Seeing positive results from Schwarz, Escolano, et al., 2020 two possibly remaining questions are:

Cross-Participant Performance

How well would the same classification setup perform in a cross-participant manner? With realistic BCI applications in mind, it may not always be affordable, or even possible, to train classification models on an individual level. This awakens interest in the decodability of new individual data. To assess this problem, a classification model was trained on the combined data of several participants and evaluated on the data of a previously unseen test participant.

Cross-System Performance

How well would the setup perform across recording systems? To investigate the feasibility of a shared classification software between electrode headsets and the transition to a new system, this question was split into two problems:

Combined training and testing systems: A classification model was trained on the combined data of two recording systems and evaluated on the combined data of the previously unseen test system.

Separate testing system: The model was trained on the combined data of 42 participants using all recording systems and evaluated using the combined, previously unseen data of the remaining 3 test participants.

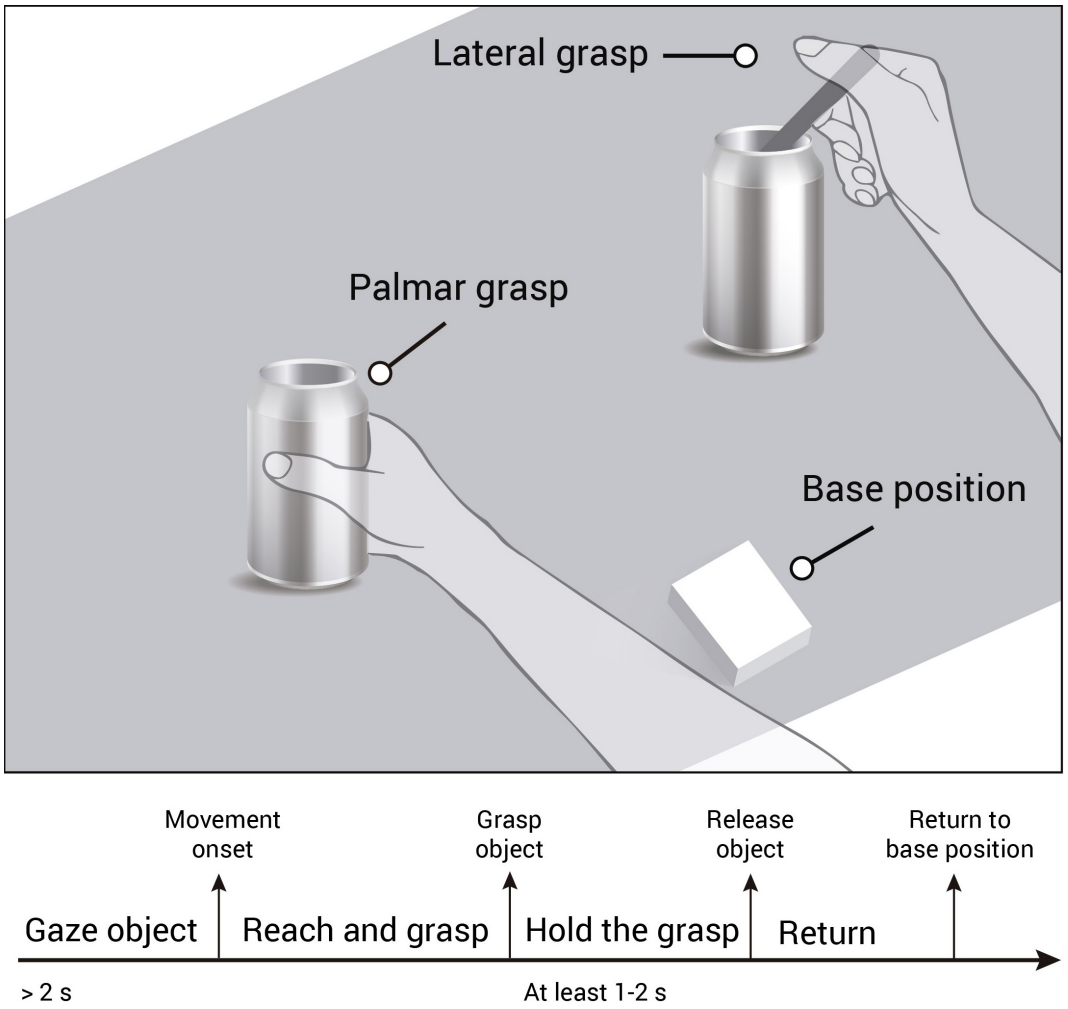

Figure 1.1: Experimental setup to record the EEG data, showing both reach and grasp actions performed by the participants (modified from Schwarz, Escolano, et al., 2020)

2 Methods

2.1 Data Preprocessing

The preprocessing and artifact handling follows closely the method described by Schwarz, Escolano, et al., 2020. The publicly available raw data was filtered using MATLAB R2021b and EEGLAB v2022.0 (Delorme and Makeig, 2004):

1. Apply a fourth order zero-phase Butterworth bandpass filter between 0.3 and 60 Hz.
2. Run the extended infomax algorithm (Lee, Girolami, and T. J. Sejnowski, 1999) for independent component analysis (ICA) using EEGLAB. Note: this step and the following was left out for the dry system "due to the unfavorable number of channels available" (Schwarz, Escolano, et al., 2020)
3. Remove independent components associated with eye artifacts by visual inspection (Delorme, Makeig, and T. Sejnowski, 2002) (see 2.5) using EEGLAB
4. Apply a fourth order zero-phase Butterworth bandpass filter between 0.3 and 35 Hz.
5. Extract trials within a window of interest (WOI) of $[-2, 3]$ seconds where the movement onset lies at second 0. Extract 81 trials of 5 seconds each from the recordings during rest condition.
6. Reject trials by statistical parameters (Faller et al., 2012; Schwarz, Scherer, et al., 2015; Schwarz, Ofner, et al., 2018) meeting one of the following conditions:
    - at least one amplitude value exceeds 125µV
    - abnormal joint probability (exceeding 4 times the standard deviation) using the EEGLAB function `jointprob`

- abnormal kurtosis (exceeding 4 times the standard deviation) using the EEGLAB function `rejkurt`.

Rejected trials were excluded from further analysis. This way, of all trials 13.84% from the gel-based sensors, 9.24% from the water-based sensors and 7.07% from the dry sensors were rejected.

7. Downsample to 16 Hz.
8. Apply a common average reference (CAR) filter using the EEGLAB function `pop_reref`.
9. Apply a fourth order zero-phase Butterworth low-pass filter with the cut-off frequency at 3Hz.

2.2 Single-Trial Multiclass Classification

To test the data for decodability the same 3 classes as by Schwarz, Escolano, et al., 2020 (rest, palmar grasp, lateral grasp) were used. The following two sections describe how the calibration and test sets were extracted from the data. The classification part stays the same for all cases:

To find the time point with the highest classification accuracy within the window of interest 80 models (16 Hz $\times$ 5 seconds WOI = 80 time points) were evaluated. For comparison a 5-fold cross-validation was implemented, which, for each time point and fold, split the calibration data into training and evaluation sets. Note that the 10 repetitions of the cross-validation approach from Schwarz, Escolano, et al., 2020 were not necessary anymore due to the higher amount of data. The actual training/testing sample for the classifier consists (for each EEG channel) of 9 amplitude values of the preceding second of the current time point, in causal steps of 0.125 seconds, the last one lying at the current time point. For clarification, the MATLAB indexing is `[-1:0.125:0]`. This yields arrays of 522 values (9 $\times$ 58 channels) for the gel-based system, 288 values (9 $\times$ 32 channels) for the water-based system and 99 values (9 $\times$ 11 channels) for the dry system.

Then, for each time point and fold, a shrinkage-based linear discriminant Bayes classifier (see section 2.6) was trained using the training set and tested on the evaluation set. The time point with the highest testing accuracy (the mean accuracy of all folds) was selected as winning time point. Using all calibration

data of this time point the winning model was trained. Thereafter, the winning model was evaluated on each time point of the previously unseen test set.

To analyze the results for statistical significance the adjusted chance level (see section 2.8) was calculated and corrected for multiple comparisons (80 time points) using Bonferroni correction.

2.3 Cross-Participant Classification

To examine the decoding performance across the data of several individuals the calibration set consisted of the combined data of 14 study participants and the test set was the data of the remaining one participant. All recordings for one classification run stem from the same system. This resulted in 15 runs per system, one for each test participant. All available EEG channels of the respective system were used (58 for the gel-based recordings, 32 for the water-based system and 11 for the dry-electrode system).

2.4 Cross-System Classification

For compatibility between recording systems only the common EEG channels could be used: [C1, C2, C3, C4, CP3, CP4, CPz, Cz, FC3, FC4, FCz]. As described in section 1.2 this task was split into two problems:

2.4.1 Training and Testing on Data of all Systems

For this problem, the combined data of 14 study participants of each recording system formed the calibration set. There were 3 testing sets per classification run, one for each recording system. Theoretically this would generate $15^3 = 3375$ possible permutations. For easier computation it was restricted to 15 classification runs, where the test participant index was kept the same for all recording systems. The low standard deviation of the results indicates no significant benefit of using more permutations. In total 45 testing results are presented (15 classification runs $\times$ 3 test participants).

2.4.2 Testing on Data of a Previously Unseen System

Here the calibration set consisted of the combined data of all 15 study participants of 2 recording systems. The data of the remaining test system formed the 15 testing sets, one for each participant.

2.5 Removing Eye Artifacts

To remove eye artifacts from the data the extended infomax algorithm was used (Lee, Girolami, and T. J. Sejnowski, 1999) for independent component analysis (ICA) decomposition. Artifacts as well as valuable information are not present at single electrodes but spread on the scalp, therefore the EEG channels are correlated. ICA takes the n-dimensional space (of the n-electrode recording) and projects it into "a more relevant coordinate frame" where "the projections of the data on each basis vector – i.e. the independent components (IC) – are maximally independent of each other" (Delorme, Makeig, and T. Sejnowski, 2002). The entries of the matrix used for backprojection are also referred to as "weights" of the ICs. The algorithm is implemented in EEGLAB as `binica` ("binary compiled ICA") with the options `("extended", 1)`.

In order to spatially analyze the independent components (e.g. by topographically plotting the component's weights on the scalp) the names of the EOG channels (as defined in the publicly available data) were mapped to the standard 10-5 system (Oostenveld and Praamstra, 2001):

- EOG-R-Top → FP2
- EOG-R-Side → AFp10h
- EOG-R-Bottom → AFp9
- EOG-L-Top → FP1
- EOG-L-Side → AFp9h
- EOG-L-Bottom → AFp10

After ICA decomposition in EEGLAB the plots of the ICs were visually inspected for eye artifacts. For comparison raw data plots of the deliberate eye blinking and eye movement periods were used. Following properties also indicate eye artifacts and were used for further confirmation of the selected artifactual

components: First, the scalp plot shows activity mainly on the frontal area (Jung et al., 2000). Secondly, the component activity power spectrum plot shows a continuous decline towards higher frequencies. And lastly, distinct eye movements and blinking in the ERP image (Makeig et al., 1999). These reasons are mentioned by Delorme, Truong, and Braboszcz, 2021 in their IC rejection tutorial. This way, for each dataset in average 3 ICs were identified as artifactual and removed from the data.

2.6 The LDA Classifier

The first step of LDA is dimensionality reduction, so if the data is projected to one dimension a classification between two classes becomes a simple threshold test, also known as Fisher's linear discriminant (Fisher, 1936). For more than two classes however, there are several options as described by Duda, 2001:

One Versus the Rest

This method leaves us with c two-class problems for c classes. For every class we assign the training samples belonging to this class in category A and all other samples in category B, which gives us c classifiers. Given a test sample we get c classification results, which must then be combined to one final result.

One Versus One

This method yields $c(c-1)/2$ binary problems for c classes. Here we train one classifier for each pair of classes. Again, we get one classification result for every two-class problem which must be combined to get a final result.

Bayes Classifier

The problem with above-mentioned methods is the ambiguity when combining results. This is illustrated for two dimensions in figure 2.1. However, this problem can be avoided by using a Bayes classifier which compares the probability of

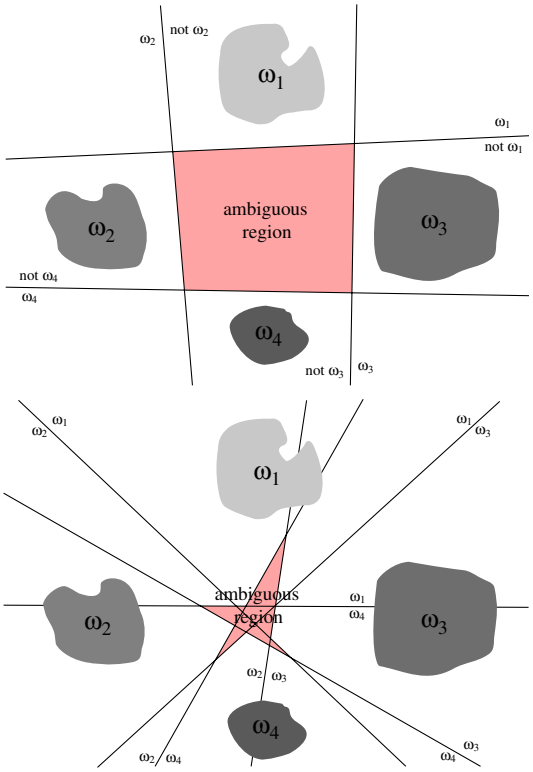

Figure 2.1: Ambiguous regions for a two-dimensional four-class problem. Top shows "one versus the rest", bottom shows "one versus one". (modified from Duda, 2001)

the sample belonging to each class. This is described in chapter 2.6.2 and 5.2.2 by Duda, 2001 and implemented in this work as follows: Let the classes ω_k ($k = 1...c$) have their individual expected sample value $\boldsymbol{\mu}_k$ as a column vector. Assuming all classes show the same covariance $\boldsymbol{\Sigma}$, calculate the weights

$$\mathbf{w}_k = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_k \tag{2.1}$$

and the bias

$$w_{k0} = -\frac{1}{2}\boldsymbol{\mu}_k^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_k. \tag{2.2}$$

The discriminant function

$$g_k(\mathbf{x}) = \mathbf{w}_k^\top \mathbf{x} + w_{k0} \tag{2.3}$$

is then a measure for the probability of the sample (feature vector) $\mathbf{x}$ belonging to class ω_k. In terms of classification, $\mathbf{x}$ is assigned to ω_k if $g_k(\mathbf{x}) > g_j(\mathbf{x})$ for all $j \neq k$.

2.7 Covariance Regularization

To estimate the covariance matrix $\mathbf{\Sigma}_k$ of the feature vectors $\mathbf{x}_1, ..., \mathbf{x}_n$ belonging to class ω_k for LDA usually the empirical covariance

$$\hat{\mathbf{\Sigma}}_k = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k)^\top \tag{2.4}$$

is used, where

$$\hat{\boldsymbol{\mu}}_k = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \tag{2.5}$$

denotes the estimated expected sample value. However, using this estimation for high-dimensional data with low sample count "leads to a systematic error: Large eigenvalues of the original covariance matrix are estimated too large, and small eigenvalues are estimated too small", which "renders LDA far from being optimal" (Blankertz et al., 2011). This can be counterbalanced by replacing $\hat{\mathbf{\Sigma}}_k$ with

$$\tilde{\mathbf{\Sigma}}_k(\gamma) = (1-\gamma)\hat{\mathbf{\Sigma}}_k + \gamma \nu_k \mathbf{I} \tag{2.6}$$

with $\gamma \in [0, 1]$, $\nu_k = \text{trace}(\hat{\mathbf{\Sigma}}_k)/d$ (average eigenvalue) and d being the dimensionality of the feature space. This method is called shrinkage or regularization (Stein, 1956). For estimation of the shrinkage parameter γ the analytic method proposed by Ledoit and Wolf, 2004 was used. In accordance to the assumption of a common covariance matrix for all classes (section 2.6) for LDA the unbiased pooled covariance matrix

$$\tilde{\mathbf{\Sigma}} = \frac{\sum_{k=1}^{3} (n_k - 1)\tilde{\mathbf{\Sigma}}_k}{\sum_{k=1}^{3} n_k - 1} \tag{2.7}$$

is used as an estimation for $\mathbf{\Sigma}$, where n_k denotes the number of samples in class k (Du et al., 2011).

2.8 Adjusting the Chance Level

The theoretical performance level of guessing the class of a given sample in a 3-class problem by chance lies at 1/3, also called *chance level*. This however is only correct for an infinite number of guesses, which obviously is not affordable. In reality there is a probability α that the real guessing performance is outside a given interval above the theoretical chance level, which follows a binomial distribution (Müller-Putz et al., 2008). To circumvent this problem, several methods exist (Genz and Kwong, 2000). Combrisson and Jerbi, 2015 provide a way to calculate the adjusted chance level in MATLAB. For n samples and c classes the probability of guessing the correct class at least z times by chance is given by:

$$P(z) = \sum_{i=z}^{n} \binom{n}{i} \cdot \left(\frac{1}{c}\right)^{i} \cdot \left(\frac{c-1}{c}\right)^{n-1}. \tag{2.8}$$

The statistically significant threshold $St(\alpha)$, or *adjusted chance level*, can then be calculated in percent using the MATLAB function `binoinv`:

$$St(\alpha) = \texttt{binoinv}(1-\alpha, n, 1/c) \cdot 100/n \tag{2.9}$$

where the significance $\alpha = \frac{z}{n}$ is the predefined acceptable probability of a false positive test.

When comparing multiple performances of the same test pipeline, the probability of one exceeding the chance level by chance of course increases with the number of tests. Bonferroni correction (Haynes, 2013) accounts for this by decreasing the significance with increasing number of comparisons m:

$$\alpha^* = \frac{\alpha}{m}. \tag{2.10}$$

In this work $\alpha = 0.05$ was corrected for comparing 80 time points and α^* was used in formula 2.9.

2.9 Data Normalization

When decoding in a cross-participant or cross-system manner (i.e. training and test set stem from different participants or recording systems) there may

be variances in setting up the EEG system on the scalp, which may result in different amplitudes for the same (real) scalp potential. The recorded amplitudes $u_i[t]$ on each electrode i are therefore normalized by the L^2 norm

$$|u_i| = \sqrt{\sum_{t=0}^{T} u_i^2[t]} \tag{2.11}$$

over the time $t = 0...T$ of the channel during rest condition.

Thereafter, the datasets are normalized by their mean global field power (GFP, Skrandies, 1990) during rest condition. The GFP at one time point given the amplitude values u_i $(i = 1...n)$ of all n electrodes can be calculated by

$$GFP = \sqrt{\frac{1}{2n} \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} (u_i - u_j)^2}. \tag{2.12}$$

Note that both normalizations are done on the preprocessed extracted trials to avoid distortion by artifact contaminated trials.

3 Results

3.1 Cross-Participant Classification

Each winning model of the 15 calibration sets was tested on the previously unseen trials of the corresponding test participant. This was performed for all time points in the window of interest. The calibration significance is calculated using formula 2.9 with $n = 14/15 \cdot n_i$, where n_i describes the total number of trials recorded by system i, after trial rejection. That leaves 3051 trials for the gel-based system, 3321 trials for the water-based system and 3390 for the dry-electrodes system. The testing significance corresponds to the remaining $n = n_i/15$. The top row of figures 3.1, 3.2 and 3.3 depicts the grand average of all 15 calibration performances. The bottom row shows the individual performances of each winning model, tested on the trials of the corresponding test participant, including the grand average thereof. The confusion matrix displays the average of the individual peak performances.

Figure 3.1 shows the results of the gel-based system. Table 3.1 depicts the corresponding accuracies for each testing set. Figure 3.2 depicts the results of the water-based system. Table 3.2 lists the accompanying values specific for the test sets. In figure 3.3 we see the results using the recordings of the dry system. Table 3.3 shows the corresponding accuracies. The standard deviation of the calibration performances between classification runs lies in average between 0.65% and 0.68% for all systems. Though plotted, it is therefore hard to see on the graphs.

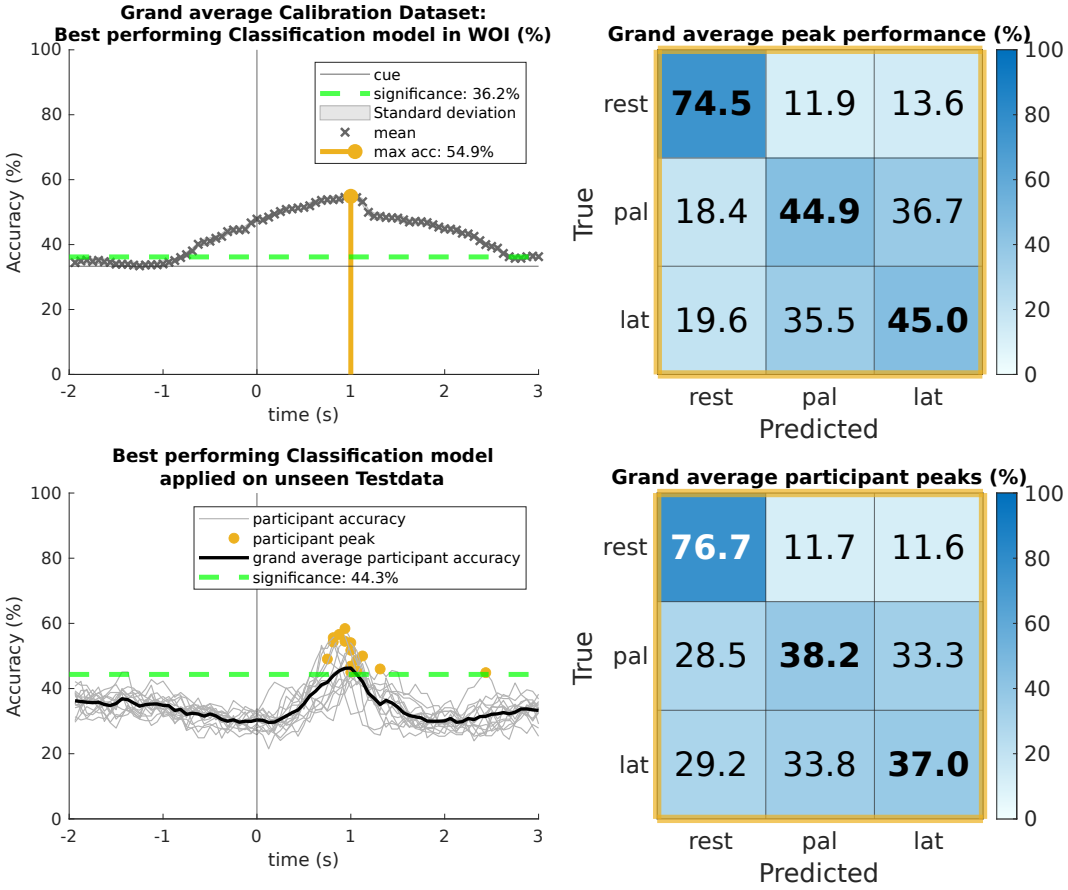

Figure 3.1: Results of the cross-participant classification using the gel-based system. Top left: Mean of all calibration accuracies during cross-validation on the whole window of interest. Top right: Row wise normalized confusion matrix corresponding to the grand average peak calibration accuracy. Bottom left: Testing accuracy on the whole window of interest of each participant, as well as the grand average. Bottom right: Row wise normalized confusion matrix corresponding to the grand average of all test-participant-specific peak performances.

Table 3.1: Cross-Participant classification results of the gel-based recordings.

Column 1 shows the test participant index. To each line belongs a classifier trained with the combined data of all participants of the gel-based system except this index. Columns 2 and 3 show peak calibration performance (%) and time of occurrence (s) with respect to the movement onset. Columns 4 and 5 show the same for the test set.

	Calibration set		**Test set**	
#	**Peak (%)**	**Time (s)**	**Peak (%)**	**Time (s)**
G01	56.3	1.0	44.9	2.4
G02	55.2	1.1	48.7	1.1
G03	56.2	1.0	45.0	1.0
G04	55.0	0.8	54.3	0.8
G05	55.8	1.0	46.9	1.0
G06	54.4	0.8	49.1	0.8
G07	55.4	0.9	51.7	1.0
G08	53.8	0.9	58.4	0.9
G09	55.5	1.0	54.1	1.0
G10	55.6	0.8	55.6	0.8
G11	55.3	1.1	50.0	1.1
G12	55.0	1.1	46.0	1.3
G13	54.6	1.0	45.4	1.1
G14	56.3	1.0	54.5	0.9
G15	54.9	0.8	56.6	0.9
Average	55.3	0.9	50.7	1.1
STD	0.7	0.1	4.6	0.4

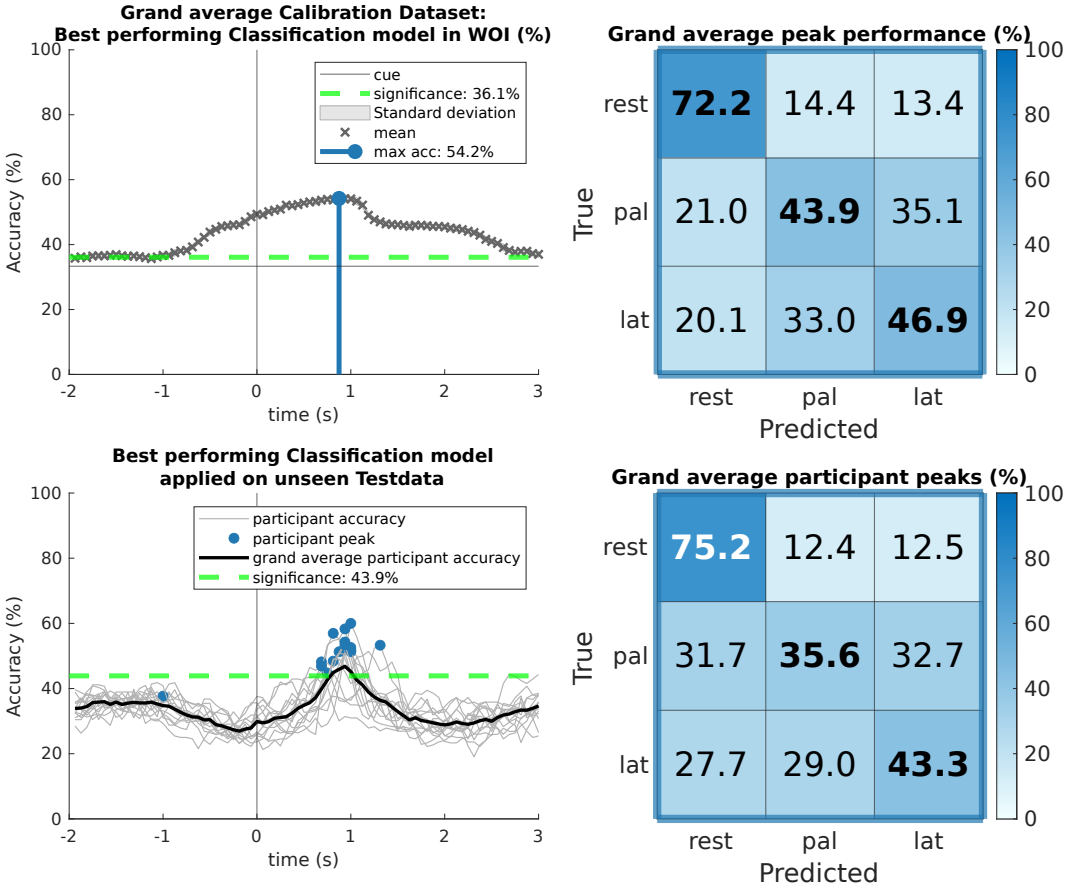

Figure 3.2: Results of the cross-participant classification using the water-based system. Top left: Mean of all calibration accuracies during cross-validation on the whole window of interest. Top right: Row wise normalized confusion matrix corresponding to the grand average peak calibration accuracy. Bottom left: Testing accuracy on the whole window of interest of each participant, including the grand average. Bottom right: Row wise normalized confusion matrix corresponding to the grand average of all test-participant-specific peak performances.

Table 3.2: Cross-Participant classification results of the water-based recordings.

Column 1 shows the test participant index. To each line belongs a classifier trained with the combined data of all participants of the water-based system except this index. Columns 2 and 3 show peak calibration performance (%) and time of occurrence (s) with respect to the movement onset. Columns 4 and 5 show the same for the test set.

	Calibration set		**Test set**	
#	**Peak (%)**	**Time (s)**	**Peak (%)**	**Time (s)**
V01	53.8	1.0	60.0	1.0
V02	53.3	0.8	56.9	0.8
V03	54.5	1.0	52.6	1.0
V04	54.0	0.9	51.3	0.9
V05	55.1	0.8	48.2	0.7
V06	54.1	0.9	58.3	0.9
V07	54.9	0.9	54.3	0.9
V08	55.2	0.9	48.4	0.8
V09	54.8	1.0	51.3	1.0
V10	54.9	1.0	51.8	1.0
V11	55.3	0.9	37.7	-1.0
V12	55.3	0.7	44.7	0.8
V13	55.3	0.7	46.9	0.7
V14	55.1	0.9	53.3	1.3
V15	54.3	0.9	53.4	0.9
Average	54.6	0.9	51.3	0.8
STD	0.6	0.1	5.6	0.5

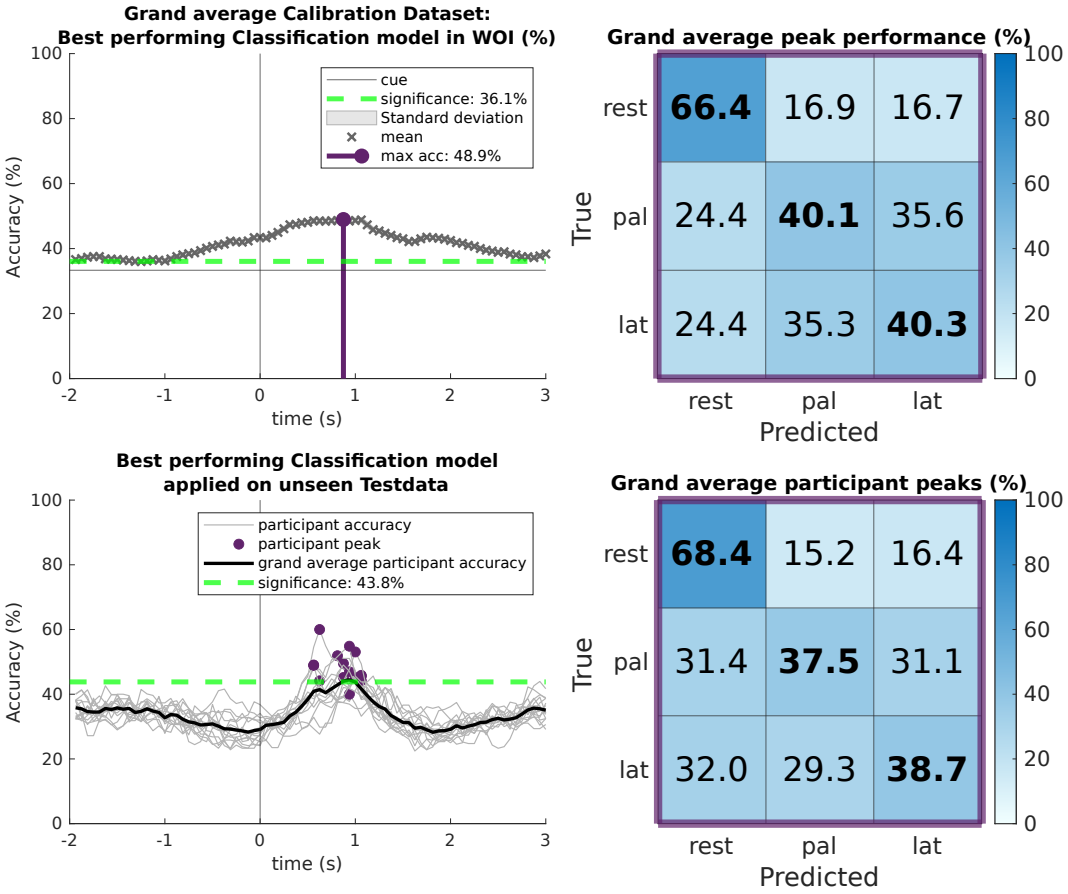

Figure 3.3: Results of the cross-participant classification using the dry system. Top left: Mean of all calibration accuracies during cross-validation on the whole window of interest. Top right: Row wise normalized confusion matrix corresponding to the grand average peak calibration accuracy. Bottom left: Testing accuracy on the whole window of interest of each participant, as well as the grand average. Bottom right: Row wise normalized confusion matrix corresponding to the grand average of all test-participant-specific peak performances.

Table 3.3: Cross-Participant classification results of the dry-electrode recordings.

Column 1 shows the test participant index. To each line belongs a classifier trained with the combined data of all participants of the dry-electrode system except this index. Columns 2 and 3 show peak calibration performance (%) and time of occurrence (s) with respect to the movement onset. Columns 4 and 5 show the same for the test set.

	Calibration set		**Test set**	
#	**Peak (%)**	**Time (s)**	**Peak (%)**	**Time (s)**
H01	49.3	1.0	45.2	0.9
H02	49.8	1.1	44.2	1.1
H03	50.3	0.6	44.1	0.6
H04	49.7	1.1	45.2	1.0
H05	49.4	0.9	49.5	0.9
H06	49.1	0.9	51.9	0.8
H07	49.2	1.0	45.8	1.1
H08	49.6	0.8	40.0	0.9
H09	48.2	0.6	60.0	0.6
H10	49.5	0.8	46.9	0.9
H11	49.5	1.1	45.6	1.1
H12	49.4	0.6	48.9	0.6
H13	49.2	1.0	53.1	1.0
H14	49.0	0.9	49.0	0.6
H15	48.9	0.9	54.8	0.9
Average	49.3	0.9	48.3	0.9
STD	0.5	0.2	5.0	0.2

3.2 Cross-System Classification

3.2.1 Training All Recording Systems

Table 3.4 shows the accuracies of the all-system classification, with both the calibration and testing data stemming from all recording systems. For each run, after calibrating the winning model using the recordings of 42 study participants across all systems, it was applied on the trials from the remaining 3 recordings.

Figure 3.4 shows the accompanying graphs to those results. The top row shows the average results of all 15 calibration sets during cross-validation. In this setup, the average standard deviation between the calibration performances of all classification runs lies at 0.96%. For the calculation of the calibration significance a sample size of $n = 42/45 \cdot \sum_i n_i$ was taken.

The bottom row shows the performance of each winning model, when applied on the previously unseen testing data from 3 participants across all systems. This results in 45 accuracy plots, excluding the additional grand average. The testing confusion matrix again corresponds to the individual best time point performances of each testing set. The testing significance stems from the remaining sample size of $n = \sum_i n_i/15$.

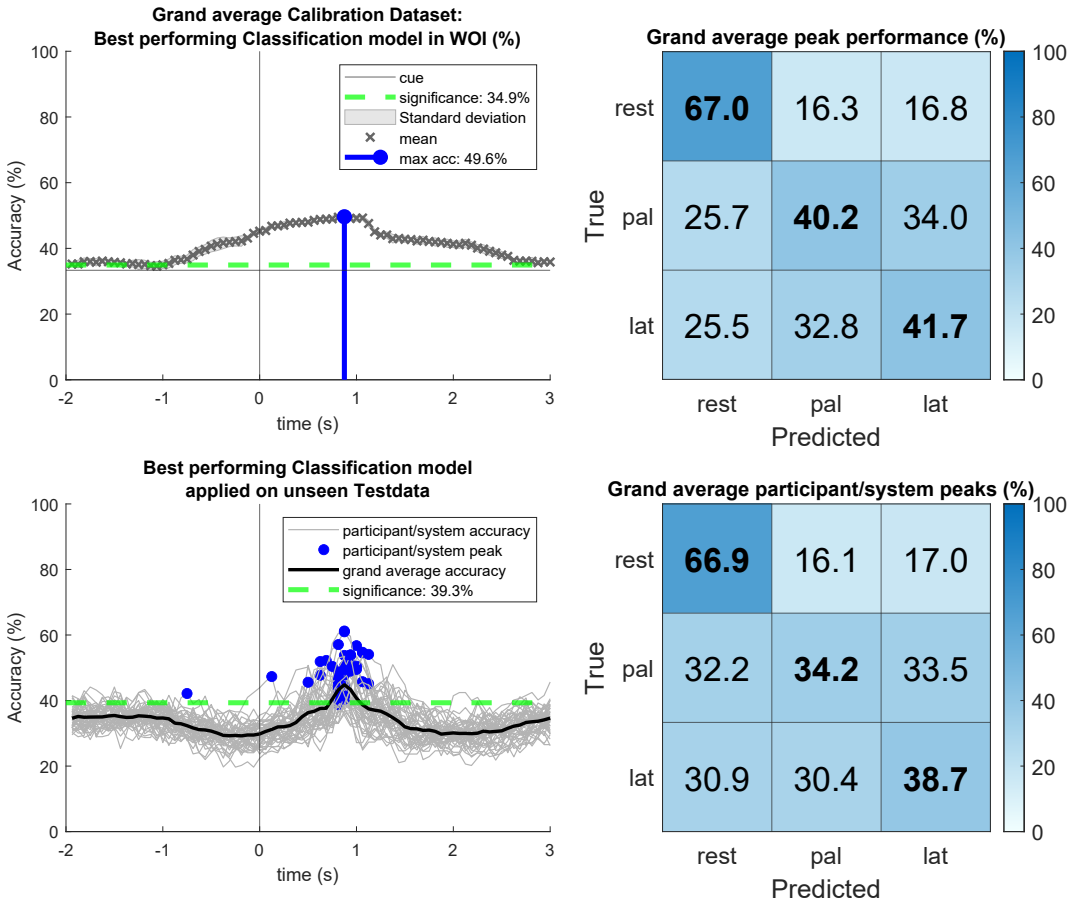

Figure 3.4: Results of the cross-system classification using all recording systems for both calibration and testing. Top left: Mean of all 15 calibration performances during cross-validation, including the time point with the highest mean accuracy. Top right: Confusion matrix corresponding to the best time point regarding average calibration accuracy. Bottom left: Performance of the winning model applied on all 45 test sets, as well as the mean thereof. Bottom right: Average confusion matrix at the individual best time points of all 45 testing sets.

Table 3.4: All-System classification results.

Column 1 shows the test participant index of the corresponding classification run. The classifier belonging to lines of the same index has been trained using the combined data of all systems and all participants except this index. Its peak calibration accuracy (%) and time of occurrence (s) with respect to the movement onset are shown in columns 2 and 3. Column 4 shows the recording system the test set stems from. Columns 5 and 6 show the peak testing accuracy (%) and time of occurrence (s).

	Calibration set		**Test set**		
#	**Peak (%)**	**Time (s)**	**System**	**Peak (%)**	**Time (s)**
01	50.7	0.9	Gel	47.3	0.1
			Versatile	50.5	0.9
			Hero	40.4	0.8
02	50.6	0.8	Gel	44.2	0.9
			Versatile	53.7	0.9
			Hero	43.3	0.8
03	49.9	0.9	Gel	41.1	0.9
			Versatile	48.7	0.8
			Hero	41.4	0.9
04	49.4	0.9	Gel	51.3	0.9
			Versatile	53.9	0.9
			Hero	40.4	0.9
05	50.9	0.8	Gel	38.8	0.8
			Versatile	47.7	0.6
			Hero	46.5	0.8
06	49.3	0.9	Gel	49.5	0.9
			Versatile	61.0	0.9
			Hero	49.4	0.9
07	49.8	0.7	Gel	52.2	0.7
			Versatile	51.9	0.6
			Hero	50.4	0.8
08	50.6	0.9	Gel	44.4	0.9
			Versatile	57.1	0.8
			Hero	41.3	0.9
09	50.0	1.1	Gel	54.5	1.1
			Versatile	45.1	1.1
			Hero	54.1	1.1

Table 3.4: All-System classification results.

Column 1 shows the test participant index of the corresponding classification run. The classifier belonging to lines of the same index has been trained using the combined data of all systems and all participants except this index. Its peak calibration accuracy (%) and time of occurrence (s) with respect to the movement onset are shown in columns 2 and 3. Column 4 shows the recording system the test set stems from. Columns 5 and 6 show the peak testing accuracy (%) and time of occurrence (s).

	Calibration set		**Test set**		
#	**Peak (%)**	**Time (s)**	**System**	**Peak (%)**	**Time (s)**
10	50.1	0.9	Gel	61.2	0.9
			Versatile	44.2	0.8
			Hero	45.6	0.8
11	50.3	0.8	Gel	50.0	0.9
			Versatile	42.2	-0.8
			Hero	47.3	0.8
12	49.8	0.9	Gel	48.4	0.9
			Versatile	47.8	0.9
			Hero	45.7	1.1
13	49.6	0.9	Gel	45.9	0.8
			Versatile	45.6	0.5
			Hero	49.0	0.9
14	49.2	1.0	Gel	56.7	1.0
			Versatile	54.7	1.1
			Hero	49.5	1.0
15	49.7	0.9	Gel	53.8	0.9
			Versatile	50.5	1.0
			Hero	53.9	0.9
Average	50.0	0.9	Gel	49.3	0.8
			Versatile	50.3	0.7
			Hero	46.6	0.9
STD	0.5	0.1	Gel	6.0	0.2
			Versatile	5.2	0.4
			Hero	4.6	0.1

3.2.2 Using a Separate Testing System

For each of the 3 classification runs, after calibrating the winning model using the combined data of two calibration systems, it was applied on each participant of the testing system. The calibration significance stems from a sample size of $n = \sum_{i,i \neq j} n_i$ with j being the testing system. The testing set has a sample size of $n = n_j/15$ (n_j ... combined number of trials recorded by the testing system), therefore the testing significance relates to the participant peak accuracies.

Figure 3.5 shows the results when applied on previously unseen recordings of the gel-based system, figure 3.6 of the water-based system and figure 3.7 of the dry system. The top row shows the calibration accuracy during cross-validation including the corresponding confusion matrix at the selected time point for the winning model. Note that, in contrast to the previous scenarios, the standard deviation is left out since there is only one classification run using the combined data of two systems. The bottom row shows the performances for each participant of the testing system, as well as the mean performance and the average confusion matrix at the participant-specific best time points.

Table 3.5 shows the corresponding test performances for each of the 45 test sets (15 participants $\times$ 3 recording systems). The calibration accuracy for the gel-based test set (training with the combined data of the water-based and dry-electrode recordings) reached its peak of 50.6% at 0.8 seconds with respect to the movement onset. For the water-based test set (training on gel-based and dry) it lies at 49.2% at 0.9 seconds. The calibration accuracy for the dry-electrode test set (training on gel-based and water-based) has its peak of 51.7% at 0.9 seconds with respect to the movement onset.

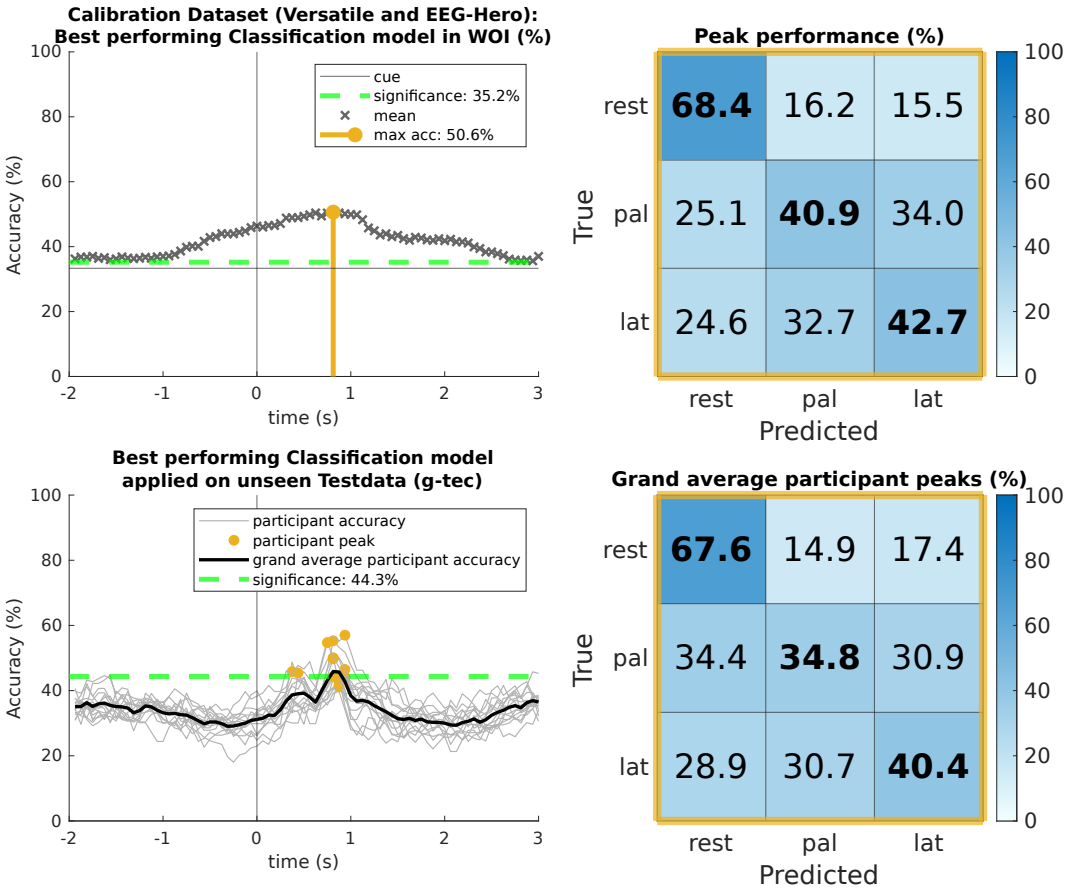

Figure 3.5: Results of the cross-system classification when testing on unseen gel-based recordings. Top left: Calibration performance during cross-validation, including the selected time point for the winning model. Top right: Confusion matrix corresponding to the selected best time point during calibration. Bottom left: Performance of the winning model applied on recordings of each participant of the previously unseen gel-based test system, including the mean thereof and the individual peak performances. Bottom right: Average confusion matrix at the individual peak performance times during testing.

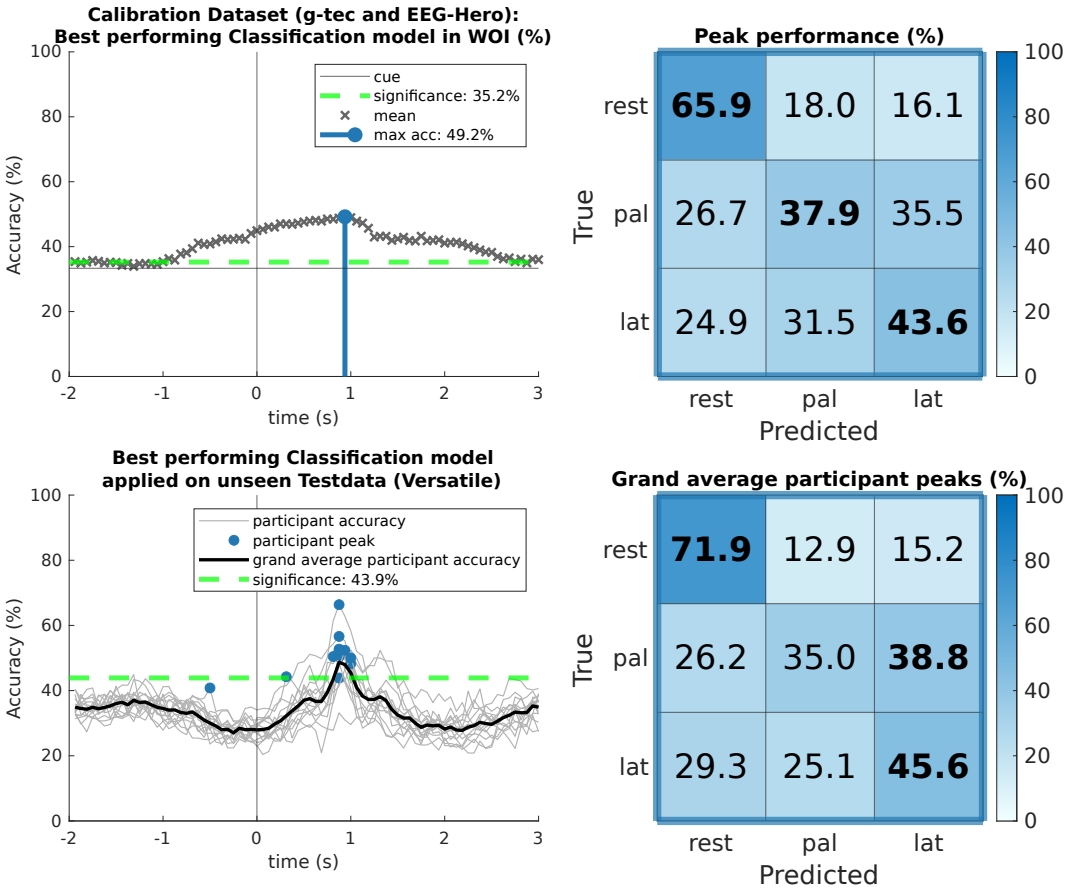

Figure 3.6: Results of the cross-system classification when testing on unseen water-based recordings. Top left: Calibration performance during cross-validation, including the selected time point for the winning model. Top right: Confusion matrix corresponding to the selected best time point during calibration. Bottom left: Performance of the winning model applied on recordings of each participant of the previously unseen water-based test system, including the mean thereof and the individual peak performances. Bottom right: Average confusion matrix at the individual peak performance times during testing.

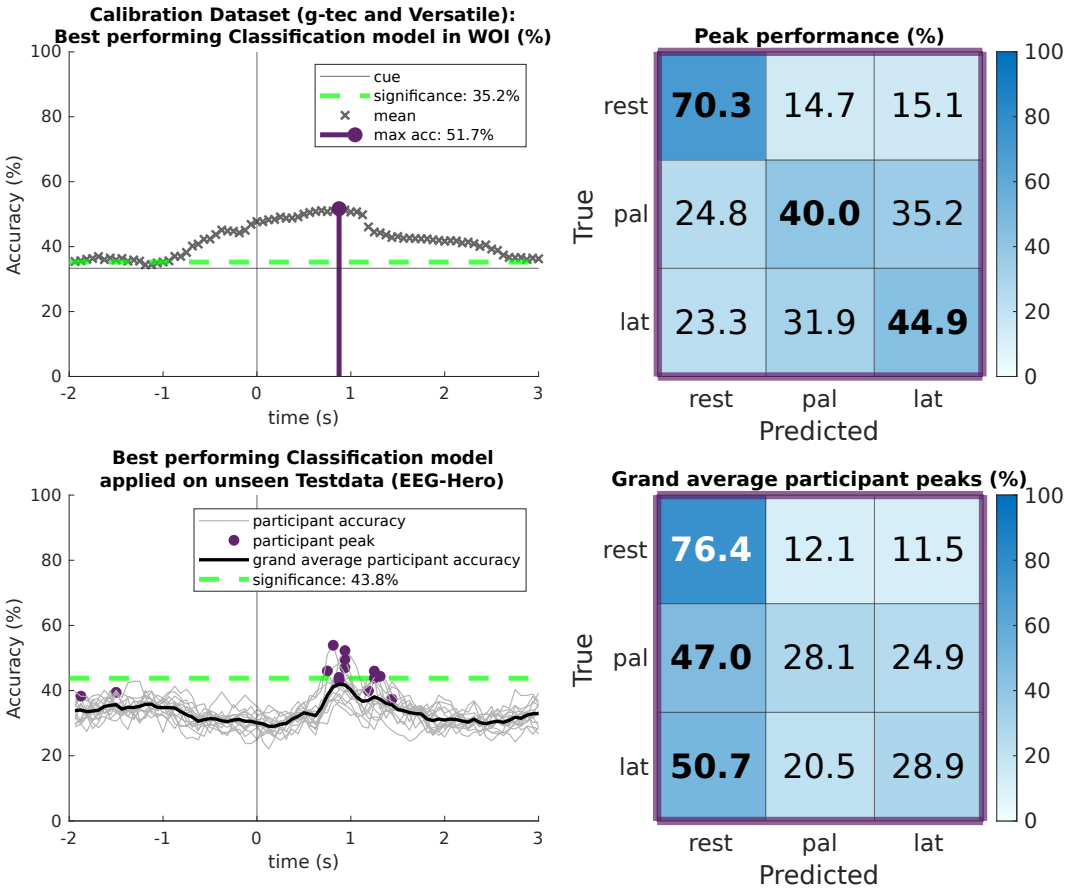

Figure 3.7: Results of the cross-system classification when testing on unseen dry-system recordings. Top left: Calibration performance during cross-validation, including the selected time point for the winning model. Top right: Confusion matrix corresponding to the selected best time point during calibration. Bottom left: Performance of the winning model applied on recordings of each participant of the previously unseen dry test system, including the mean thereof and the individual peak performances. Bottom right: Average confusion matrix at the individual peak performance times during testing.

Table 3.5: Classification results of all test systems when trained on the other two systems.

The headings "Gel" "Water" and "Dry" describe the recording system used for the test set. Those columns belong to classifiers trained with the combined data of the two remaining systems. Column 1 shows the test participant index. Columns 2, 4 and 6 show the peak performances (%) of the test set and columns 3, 5 and 7 the time of occurrence (s) with respect to the movement onset for each test set.

	Gel		**Water**		**Dry**	
#	**Peak (%)**	**Time (s)**	**Peak (%)**	**Time (s)**	**Peak (%)**	**Time (s)**
01	43.9	0.9	52.7	0.9	39.5	-1.5
02	44.2	0.9	50.9	0.9	39.9	1.2
03	41.1	0.9	52.2	0.9	37.4	1.4
04	55.3	0.8	50.4	0.9	43.0	0.9
05	44.5	0.9	50.5	0.9	47.0	0.9
06	50.0	0.8	66.4	0.9	46.0	1.2
07	42.3	0.9	51.0	0.9	49.6	0.9
08	45.8	0.4	56.6	0.9	38.3	-1.9
09	49.8	0.8	48.2	1.0	52.3	0.9
10	57.0	0.9	44.2	0.3	43.9	1.2
11	45.4	0.4	40.8	-0.5	46.0	0.8
12	46.5	0.9	43.9	0.9	44.4	1.3
13	44.3	0.8	50.4	0.8	44.1	0.9
14	50.0	0.8	52.4	0.9	43.2	0.9
15	54.7	0.8	50.0	1.0	53.9	0.8
Average	47.6	0.8	50.7	0.8	44.6	0.7
STD	4.9	0.2	5.9	0.4	4.8	1.0

4 Discussion

4.1 Confirmations

The presented classification approach followed closely the steps described by Schwarz, Escolano, et al., 2020. After trial extraction 80 classification models within a time window of interest were compared regarding their accuracy. The winning model was then tested on a previously unseen testing set. This was done using various selections of training and testing sets across study participants and recording systems.

The cross-participant same-system classification results show a decodability of rest-vs-movement, although the false positives (falsely classified as rest) increase in the testing step. Movement-vs-movement decodability is indicated but never above the adjusted chance level. The gel and water-based electrodes recordings show comparable results, for the dry system there is an overall decrease in performance. This was expected and attributed to the decreased number of channels available, as it is already mentioned by Schwarz, Escolano, et al., 2020 and Schwarz, Ofner, et al., 2018. The movement-vs-rest accuracy indicates common properties in the recordings regarding movement-related cortical potentials between individuals. The decodability is likely attributed to the "Bereitschaftspotential" (Shibasaki and Hallett, 2006).

When training and testing using the data of all systems the rest condition stays distinguishable. However, the false positives increase heavily on the test set. When given a movement trial the classifier spreads its predictions almost evenly with movement-vs-movement accuracy merely above theoretical chance level.

After training on the data of 2 recording systems and testing on the third system, the movement-vs-rest accuracy behaves very similar to above-mentioned

scenarios. Especially when testing with the gel-based recordings movement-vs-movement decodability is indicated, although below statistical significance. When evaluating the water-based system recordings there seems to be a prediction bias towards lateral grasp condition. When testing the dry electrode recordings a heavy bias appears on the rest condition, which likely is related to the two outliers of peak test accuracy regarding the time of occurrence. These results imply that when comparing recordings of different electrode types significant changes in the scalp potential (like the "Bereitschaftspotential" which indicates an oncoming movement) stay recognizable. However, finer differences between conditions seem to get lost.

To answer the questions from section 1.2: The overall performances of all scenarios are significantly higher than chance level, whereas the two movement classes could never be differentiated with a significant accuracy.

4.2 Limitations

The negative results in movement-vs-movement decodability are restricted to the documented classification pipeline and do not necessarily mean no decodability at all. They may be caused by several limitations of this work:

All steps except the artifact removal through independent component rejection are completely reproducible. This is because identifying artifacts "by visual inspection", as described by Schwarz, Escolano, et al., 2020, is a subjective task. To circumvent this, the EEGLAB toolbox provides automatic labeling of components which, however, was only used for assistance and not for rigorous component rejection in this work.

The presented classifier bases upon linear discriminant analysis, a dimensionality reduction method and rather conservative and simple technique for classification. By now there are more sophisticated approaches like support vector machines (Cortes and Vapnik, 1995) or deep learning through artificial neural networks (Craik, He, and Contreras-Vidal, 2019). Maybe even principal component analysis (PCA) could be a candidate, as it sometimes performs better according to Martinez and Kak, 2001.

For faster computation the sampling rate of the recordings was reduced to 16 Hz and only 9 amplitude values during the preceding second of the inspected time point were used for classification. This simplification could drop important information and has to be examined using more computational power than the GPU of a common desktop computer.

To find statistically significant performance values in the whole window of interest Bonferroni correction was used. This is a conservative method, other approaches (Glickman, Rao, and Schultz, 2014, Chen, Feng, and Yi, 2017) might find significance were Bonferroni correction fails.

4.3 Remarks

Since the publicly available datasets are stored in single (32 bit) floating point precision but scientific computations are often done in double (64 bit) precision (e.g. MATLAB's default datatype is `double`), both types were tried and results showed no significant differences. The presented results were computed using single precision.

For comparability with the related work of Schwarz, Escolano, et al., 2020, their classification results were reproduced beforehand. Then only the necessary parts of the MATLAB code were changed to produce the presented results. The design of the plots and tables is also inspired by their work.

The analytical method for estimating the covariance matrix shrinkage parameter (see section 2.7) did not always produce the best results. Often, a shrinkage parameter $\gamma = 0.3$ or $\gamma = 0.4$ preformed slightly better. This may be due to the high amount of training samples available. However, according to scientific practice regarding hypothesis testing, parameters could not be changed after seeing the results. The presented results are all achieved using the analytical method.

5 Conclusion

This work presents a single-trial multiclass based decoding approach using EEG recordings which include both movement conditions and rest. The recordings originate from systems with different electrode types and include several study participants for each system. It could be shown that for each scenario, incorporating comprehensive decoding across participants and recording systems, the rest condition was always identifiable against movement conditions with a low false negative rate, but a quite high false positive rate. However, with this rather conservative linear discriminant analysis setup, the movement-vs-movement performance was never statistically significant. For the interested reader the MATLAB code used for this work has been published on GitHub[1].

[1] https://github.com/mhastu/bac

Bibliography

Blankertz, Benjamin et al. (2011). "Single-trial analysis and classification of ERP components — A tutorial." In: *NeuroImage* 56.2. Multivariate Decoding and Brain Reading, pp. 814–825. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2010.06.048. URL: https://www.sciencedirect.com/science/article/pii/S1053811910009067 (cit. on p. 10).

Chen, Shi-Yi, Zhe Feng, and Xiaolian Yi (2017). "A general introduction to adjustment for multiple comparisons." In: *Journal of Thoracic Disease* 9.6. ISSN: 2077-6624. URL: https://jtd.amegroups.com/article/view/13609 (cit. on p. 31).

Combrisson, Etienne and Karim Jerbi (July 2015). "Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy." In: *Journal of Neuroscience Methods* 250, pp. 126–136. DOI: 10.1016/j.jneumeth.2015.01.010. URL: https://doi.org/10.1016%2Fj.jneumeth.2015.01.010 (cit. on p. 11).

Cortes, Corinna and Vladimir Vapnik (Sept. 1995). "Support-vector networks." In: *Machine Learning* 20.3, pp. 273–297. DOI: 10.1007/bf00994018. URL: https://doi.org/10.1007/bf00994018 (cit. on p. 30).

Craik, Alexander, Yongtian He, and Jose L Contreras-Vidal (Apr. 2019). "Deep learning for electroencephalogram (EEG) classification tasks: a review." In: *Journal of Neural Engineering* 16.3, p. 031001. DOI: 10.1088/1741-2552/ab0ab5. URL: https://doi.org/10.1088/1741-2552/ab0ab5 (cit. on p. 30).

Delorme, Arnaud and Scott Makeig (2004). "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis." In: *Journal of Neuroscience Methods* 134.1, pp. 9–21. ISSN: 0165-0270. DOI: https://doi.org/10.1016/j.jneumeth.2003.10.009. URL: https://www.sciencedirect.com/science/article/pii/S0165027003003479 (cit. on p. 4).

Delorme, Arnaud, Scott Makeig, and Terrence Sejnowski (Jan. 2002). "Automatic Artifact Rejection For EEG Data Using High-Order Statistics And Independent Component Analysis." In: 7 (cit. on pp. 4, 7).

Delorme, Arnaud, Dung Truong, and Claire Braboszcz (2021). *Independent Component Analysis for artifact removal*. commit 9f8c0122a9b23fae1043fc69c0aea5cd69a2991e. URL: https://github.com/sccn/sccn.github.io/blob/master/tutorials/06_RejectArtifacts/RunICA.md (visited on 09/29/2022) (cit. on p. 8).

Du, Jiang et al. (2011). "Automated vessel segmentation using cross-correlation and pooled covariance matrix analysis." eng. In: *Magnetic resonance imaging* 29.3, pp. 391–400. ISSN: 0730-725X (cit. on p. 10).

Duda, Richard O (2001). *Pattern classification*. eng. 2. ed.. A Wiley-Interscience publication. ISBN: 9780471056690 (cit. on pp. 8, 9).

Faller, Josef et al. (2012). "Autocalibration and Recurrent Adaptation: Towards a Plug and Play Online ERD-BCI." In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 20.3, pp. 313–319. DOI: 10.1109/TNSRE.2012.2189584 (cit. on p. 4).

Fisher, R. A. (1936). "The use of multiple measurements in taxonomic problems." In: *Annals of Eugenics* 7.2, pp. 179–188. DOI: https://doi.org/10.1111/j.1469-1809.1936.tb02137.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-1809.1936.tb02137.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-1809.1936.tb02137.x (cit. on p. 8).

Genz, Alan and Koon-Shlng. Kwong (2000). "Numerical evaluation of singular multivariate normal distributions." In: *Journal of Statistical Computation and Simulation* 68.1, pp. 1–21. DOI: 10.1080/00949650008812053. eprint: https://doi.org/10.1080/00949650008812053. URL: https://doi.org/10.1080/00949650008812053 (cit. on p. 11).

Glickman, Mark E., Sowmya R. Rao, and Mark R. Schultz (2014). "False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies." In: *Journal of Clinical Epidemiology* 67.8, pp. 850–857. ISSN: 0895-4356. DOI: https://doi.org/10.1016/j.jclinepi.2014.03.012. URL: https://www.sciencedirect.com/science/article/pii/S0895435614001127 (cit. on p. 31).

Haynes, Winston (2013). "Bonferroni Correction." In: *Encyclopedia of Systems Biology*. Ed. by Werner Dubitzky et al. New York, NY: Springer New York, pp. 154–154. ISBN: 978-1-4419-9863-7. DOI: 10.1007/978-1-4419-9863-

7_1213. URL: https://doi.org/10.1007/978-1-4419-9863-7_1213 (cit. on p. 11).

Jung, Tzyy-Ping et al. (2000). "Removing electroencephalographic artifacts by blind source separation." In: *Psychophysiology* 37.2, pp. 163–178. DOI: https://doi.org/10.1111/1469-8986.3720163. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1469-8986.3720163. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1469-8986.3720163 (cit. on p. 8).

Ledoit, Olivier and Michael Wolf (2004). "A well-conditioned estimator for large-dimensional covariance matrices." In: *Journal of Multivariate Analysis* 88.2, pp. 365–411. ISSN: 0047-259X. DOI: https://doi.org/10.1016/S0047-259X(03)00096-4. URL: https://www.sciencedirect.com/science/article/pii/S0047259X03000964 (cit. on p. 10).

Lee, Te-Won, Mark Girolami, and Terrence J. Sejnowski (Feb. 1999). "Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources." In: *Neural Computation* 11.2, pp. 417–441. ISSN: 0899-7667. DOI: 10.1162/089976699300016719. eprint: https://direct.mit.edu/neco/article-pdf/11/2/417/814042/089976699300016719.pdf. URL: https://doi.org/10.1162/089976699300016719 (cit. on pp. 4, 7).

López-Larraz, Eduardo et al. (2014). "Continuous decoding of movement intention of upper limb self-initiated analytic movements from pre-movement EEG correlates." In: *Journal of NeuroEngineering and Rehabilitation* 11.1, p. 153. DOI: 10.1186/1743-0003-11-153. URL: https://doi.org/10.1186%2F1743-0003-11-153 (cit. on p. 1).

Makeig, Scott et al. (1999). "Functionally Independent Components of the Late Positive Event-Related Potential during Visual Spatial Attention." In: *Journal of Neuroscience* 19.7, pp. 2665–2680. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.19-07-02665.1999. eprint: https://www.jneurosci.org/content/19/7/2665.full.pdf. URL: https://www.jneurosci.org/content/19/7/2665 (cit. on p. 8).

Martinez, A.M. and A.C. Kak (2001). "PCA versus LDA." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23.2, pp. 228–233. DOI: 10.1109/34.908974 (cit. on p. 30).

Müller-Putz, Gernot et al. (Jan. 2008). "Better than Random? A closer look on BCI results." In: *International Journal of Bioelektromagnetism* 10, pp. 52–55 (cit. on p. 11).

Oostenveld, Robert and Peter Praamstra (2001). "The five percent electrode system for high-resolution EEG and ERP measurements." In: *Clinical Neurophysiology* 112.4, pp. 713–719. ISSN: 1388-2457. DOI: https://doi.org/10.1016/S1388-2457(00)00527-7. URL: https://www.sciencedirect.com/science/article/pii/S1388245700005277 (cit. on p. 7).

Pei, Dingyi et al. (2020). "Neural Decoding of Upper Limb Movements Using Electroencephalography." In: *Brain–Computer Interface Research: A State-of-the-Art Summary 8*. Ed. by Christoph Guger, Brendan Z. Allison, and Kai Miller. Cham: Springer International Publishing, pp. 25–33. ISBN: 978-3-030-49583-1. DOI: 10.1007/978-3-030-49583-1_3. URL: https://doi.org/10.1007/978-3-030-49583-1_3 (cit. on p. 1).

Schwarz, Andreas, Carlos Escolano, et al. (2020). "Analyzing and Decoding Natural Reach-and-Grasp Actions Using Gel, Water and Dry EEG Systems." In: *Frontiers in Neuroscience* 14. ISSN: 1662-453X. DOI: 10.3389/fnins.2020.00849. URL: https://www.frontiersin.org/articles/10.3389/fnins.2020.00849 (cit. on pp. 1–5, 29–31).

Schwarz, Andreas, Patrick Ofner, et al. (Feb. 2018). "Decoding natural reach-and-grasp actions from human EEG." In: *J. Neural Eng.* 15.1, p. 016005. DOI: 10.1088/1741-2552/aa8911. URL: https://doi.org/10.1088/1741-2552/aa8911 (cit. on pp. 4, 29).

Schwarz, Andreas, Reinhold Scherer, et al. (Aug. 2015). "A co-adaptive sensory motor rhythms Brain-Computer Interface based on common spatial patterns and Random Forest." In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. DOI: 10.1109/embc.2015.7318545. URL: https://doi.org/10.1109%2Fembc.2015.7318545 (cit. on p. 4).

Shibasaki, Hiroshi and Mark Hallett (2006). "What is the Bereitschaftspotential?" In: *Clinical Neurophysiology* 117.11, pp. 2341–2356. ISSN: 1388-2457. DOI: https://doi.org/10.1016/j.clinph.2006.04.025. URL: https://www.sciencedirect.com/science/article/pii/S138824570600229X (cit. on p. 29).

Skrandies, Wolfgang (1990). "Global field power and topographic similarity." In: *Brain Topography* 3.1, pp. 137–141. DOI: 10.1007/bf01128870. URL: https://doi.org/10.1007%2Fbf01128870 (cit. on p. 12).

Stein, Charles M. (1956). "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution." In: (cit. on p. 10).

Suwannarat, Arpa, Setha Pan-Ngum, and Pasin Israsena (Aug. 2018). "Comparison of EEG measurement of upper limb movement in motor imagery training system." en. In: *Biomed. Eng. Online* 17.1, p. 103. DOI: 10.1186/s12938-018-0534-0. URL: https://doi.org/10.1186/s12938-018-0534-0 (cit. on p. 1).

Úbeda, Andrés et al. (Feb. 2017). "Classification of upper limb center-out reaching tasks by means of EEG-based continuous decoding techniques." In: *Journal of NeuroEngineering and Rehabilitation* 14.1. DOI: 10.1186/s12984-017-0219-0. URL: https://doi.org/10.1186%2Fs12984-017-0219-0 (cit. on p. 1).

Wolpaw, Jonathan R et al. (June 2002). "Brain–computer interfaces for communication and control." In: *Clinical Neurophysiology* 113.6, pp. 767–791. DOI: 10.1016/s1388-2457(02)00057-3. URL: https://doi.org/10.1016%2Fs1388-2457%2802%2900057-3 (cit. on p. 1).