Machine Learning Methods for Brain State-Dependent Motor Rehabilitation After Stroke

---

Dissertation zur Erlangung des Doktorgrades der
Technischen Fakultät der
Albert-Ludwigs-Universität Freiburg im Breisgau

---

vorgelegt von

Andreas Meinel
geb. in Donaueschingen

DEKANIN:
Prof. Dr. Hannah Bast

PRÜFUNGSKOMMISSION:

Prof. Dr. Bernhard Nebel (Vorsitzender)
Albert-Ludwigs-Universität Freiburg

Dr. Michael Tangermann (Erstgutachter und Betreuer)
Albert-Ludwigs-Universität Freiburg

Prof. Dr. Fabien Lotte (Zweitgutachter)
INRIA Bordeaux Sud-Ouest/LaBRI

Jun.-Prof. Dr. Joschka Bödecker (Drittgutachter)
Albert-Ludwigs-Universität Freiburg

Prof. Dr. Frank Hutter (Beisitzer)
Albert-Ludwigs-Universität Freiburg

DATUM DER DISPUTATION:
04. Juni 2019

ABSTRACT

Progressively, machine learning (ML) methods proved useful in various biomedical applications, such as brain-computer interfaces (BCIs). Such systems utilize ongoing brain signals by either translating them into messages or commands for an application, or by adapting a system depending on the user's brain state. As BCIs allow to directly interact with the human brain in real-time, they are increasingly deployed for clinical applications, e.g. in post-stroke rehabilitation. Here, ML-based brain state decoding approaches are indispensable building blocks to extract user-specific information from high-dimensional brain activity recordings with an inherently low signal-to-noise ratio on single-trial level. However, in patient scenarios current decoding methods are stretched to their limitations. The main drawbacks are their limited robustness with scarce training data, the strong hyperparameter sensitivity of trained models as well as the limitation of the feature's clinical interpretability.

To tackle the aforementioned challenges, this interdisciplinary thesis holds ML-based contributions and strikes a novel closed-loop application path.

First, novel algorithmic contributions for robust and functionally relevant brain state decoding are developed. Focusing on a supervised multivariate spatial filtering algorithm to decode a continuous variable from brain signals, different regularization strategies are introduced and evaluated on simulation and real-world datasets. Overall, the novel regularized methods reveal increased decoding performance up to 27 % under scarce training data and label noise conditions. To additionally capture the functional role of features, a novel algorithmic strategy is provided which exploits the within-trial structure of tasks. Therefore, the observed model variability (e.g. by varying the hyperparameters) is harnessed by condensing the functional signatures of a large set of models into homogeneous clusters of individual oscillatory brain dynamics.

Second, the developed ML tools are directly transferred into a closed-loop system. Focusing on post-stroke rehabilitation, data of a repetitive motor task is collected and studied in which large trial-to-trial performance variations occur. With the ML tools at hand, this variability can partly be explained and even temporally predicted by assessing individual oscillatory brain activity—shown on data of normally aged controls and stroke patients. As a novelty, individualized predictors were exploited in a proof-of-concept study on a brain state-dependent temporal gating strategy with four chronic stroke patients. In contrast to the direct decoding of user intention, the ongoing brain state revealed favorable starting time points to influence upcoming single-trial motor performance, such as reaction time (RT). In this way, RT was significantly shortened under *suitable* brain states by up to 40 % of individual RT fluctuations. Gaining performance

influence by exploiting the ongoing brain state holds a promising value for motor learning after stroke and could potentially be applied to optimize performance in cognitive trainings or sports science.

Overall, the provided ML tools foster the data-driven characterization of neural processes in complex tasks and help to deepen the global understanding of motor control mechanisms and their restoration after stroke.

ZUSAMMENFASSUNG

Methoden aus dem Bereich des maschinellen Lernens (ML) haben sich zunehmend in verschiedenen biomedizinischen Anwendungen als äußerst nützlich erwiesen. Ein Beispiel dafür sind Gehirn-Computer Schnittstellen (engl.: *Brain-Computer Interfaces*, BCIs). Solche BCI Systeme übersetzen Gehirnsignale entweder in Nachrichten oder Steuerungssignale einer Anwendung oder adaptieren ein System in Abhängigkeit des aktuellen Gehirnzustandes. Da BCI-Systeme die direkte Interaktion mit dem menschlichen Gehirn in Echtzeit ermöglichen, werden diese Systeme vermehrt für klinische Anwendungen, wie zum Beispiel Schlaganfall-Rehabilitation, eingesetzt. Hierfür sind ML-basierte Methoden zur Dekodierung von Gehirnzuständen unverzichtbare Bestandteile, um individuelle Informationen aus hochdimensionalen Gehirnaktivitätsmessungen mit intrinsisch geringem Signal-Rausch-Verhältnis auf Einzeltrial Basis zu gewinnen. Aktuelle Dekodierungsansätze, die speziell für klinische Patientenszenarien geeignet sind, stoßen jedoch rasch an ihre Grenzen. Hierzu zählen die begrenzte Robustheit der Methoden im Falle kaum verfügbarer Trainingsdaten, die starke Hyperparameter-Sensitivität der trainierten ML-Modelle sowie die begrenzte klinische Interpretierbarkeit der zugehörigen Merkmale.

Um diesen Herausforderungen zu begegnen, beinhaltet diese interdisziplinäre Dissertation neue ML-basierte Methoden und schlägt einen neuen Weg für eine closed-loop Anwendung ein.

Zunächst werden neue algorithmische Beiträge zur robusten Dekodierung von funktionell relevanten Gehirnzuständen entwickelt. Hierzu werden für ein überwachtes Lernverfahren zur Dekodierung einer kontinuierlichen Variablen aus Gehirnsignalen verschiedene Regularisierungsstrategien eingeführt. Die regularisierten, multivariaten Methoden werden auf simulierten und echten Daten evaluiert. Stehen nur wenige Trainingsdaten mit verrauschten Labels zur Verfügung, zeigen die neuen regularisierten Varianten insgesamt eine verbesserte Dekodierungsgüte von bis zu 27 %. Um zusätzlich die funktionale Rolle von Merkmalen zu erfassen, wird ein neuer algorithmischer Ansatz vorgestellt, der die innere Trial-Struktur von Paradigmen ausschöpft. Dazu wird die beobachtete Modell-Variabilität (z. B. unter verschiedenen Hyperparametern) genutzt, indem die funktionale Signatur einer umfassenden Menge an Modellen in homogene Cluster individueller oszillatorischer Gehirnaktivität kondensiert wird.

Die entwickelten ML-basierten Werkzeuge werden direkt in ein neues closed-loop System transferiert. Im Kontext Schlaganfall-Rehabilitation wurden Daten von gesunden Kontrollen und chronischen Schlaganfall-Patienten einer wiederholten Bewegungsaufgabe aufgenommen und analysiert. Hier fällt eine deutliche Leistungsvariabilität auf. Basiernd auf individuellen oszillatorischen Signalen konnten die entwickelten ML-Methoden diese Variabilität teilweise erklären und sogar zeitlich vorhersagen. In einer

Machbarkeitsstudie mit vier chronischen Schlaganfall-Patienten wurden die individuell bestimmten Prädiktoren für ein Gehirnzustand-basiertes Training eingesetzt. Im Gegensatz zur herkömmlichen Dekodierung der Nutzerintention, wurden hier erstmals Gehirnzustände dekodiert, um geeignete Startzeitpunkte einer bevorstehenden Bewegungsaufgabe auf Einzeltrial Ebene vorherzusagen und zeitlich auszulösen. Dabei wurden Reaktionszeiten (RT) unter *günstigen* Gehirnzuständen signifikant um bis zu 40 % bezüglich der auftretenden, individuellen RT-Fluktuationen verkürzt. Einfluss auf die (Bewegungs-)Qualität durch die Ausnutzung des vorherrschenden Gehirnzustands zu gewinnen hat vielversprechenden Wert für motorisches Lernen nach Schlaganfall aber auch für andere Felder, in denen Leistung optimiert wird. Dies könnte zum Beispiel für kognitive Trainingsanwendungen oder in den Sportwissenschaften der Fall sein.

Insgesamt fördern die bereitgestellten ML-basierten Methoden die datengetriebene Charakterisierung von neuronalen Prozessen in komplexen Bewegungsaufgaben. Sie helfen damit das globale Verständnis von Motoransteuerungs-Mechanismen und deren Wiederherstellung nach Schlaganfall zu vertiefen.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

The most frequently used variables as well as standardized methods are listed below. Please note that variables are often utilized with additional subscript indices and will then be explained in the corresponding text section. Hereafter, the specified page link refers to the first appearance of the variable or method in this thesis. Also note that chapters, sections, equations, figures and tables are concisely referenced in text as Chap., Sec., Eq., Fig. and Tab., respectively.

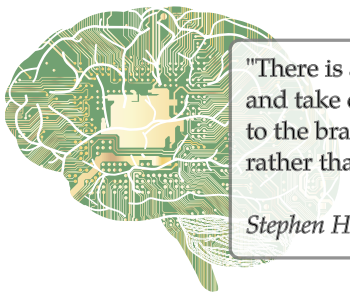
NOMENCLATURE

Part I

INTRODUCTION & FUNDAMENTALS

The first part of this thesis is dedicated to an overall introduction, the formulation of the tackled research questions and an overview of the fundamental methods and findings on which this dissertation is directly based on.

1 INTRODUCTION



"There is a real danger that computers will develop intelligence and take over. We urgently need to develop direct connections to the brain so that computers can add to human intelligence rather than be in opposition."

Stephen Hawking (2001)

Progressively, machine learning (ML) methods have proved useful in various biomedical applications, driven by the increased amount of available digitized data [1]. Targeted scenarios cover a broad scope ranging from large-scale annotation of genome sequences [2], via the support of clinical decision making, e.g., on cancer diagnosis [3, 4] through to real-time cardiac monitoring to recognize abnormalities [5]. In all these applications, the studied complex biological systems typically provide high-dimensional and noisy data. Thus, inferring variables or states of interest from such data requires coping with these characteristics by providing robust and reliable ML models. This also includes an adequate validation and understanding of the trained models [6].

A prominent neurotechnological system that involves machine learning algorithms is a brain-computer interface (BCI) [7]. Such systems allow their users to interact with a computer or physical device through their brain activity [8]. The extensive use of machine learning algorithms in BCI applications enable for single-trial decoding from high-dimensional brain activity recordings which are subject to an inherently low signal-to-noise ratio [9]. BCIs allow for closed-loop applications and are originally applied for communication and control systems.

During the last decade, BCI methods were increasingly utilized in post-stroke motor rehabilitation, with a strong emphasis on upper limb systems [10–13]. Here, BCIs enable to directly decode movement related information on single-trial level to trigger targeted and specific feedback [14], such as the movement of an orthosis [15] or the onset of an electrical stimulation [16]. Ideally, such direct feedback fosters neuroplasiticity to bypass interrupted neural pathways. Up to now, first promising results from randomized controlled studies on BCI-supported rehabilitation trainings were reported. Due to the limited amount of studies so far, an overall conclusion about their efficiency cannot be drawn up to now [17]. In such clinical applications, an indispensable building block is the availability of ML-based decoding methods that provide robust and functionally relevant features across multiple sessions. To reach this goal, Makeig et al.

[18] recommended to put strong emphasis on model interpretability for future decoding approaches and to directly incorporate neurophysiological information into data-driven models.

Broadening the view on cognitive or mental state assessment [19], Jensen et al. [20] proposed the general concept of brain state-dependent (BSD) experimenting to specifically exploit the ongoing brain state of a user in closed-loop applications. In contrast to the direct decoding of motor tasks, the BSD concept enlarges the focus by monitoring and exploiting additional cognitive processes. Examples herefore are related to perception and attention in the visual, auditory and somatosensory domain. However, up to now only a limited amount of online studies are following this concept [21].

In this thesis the recently proposed BSD concept will be established for a repetitive hand motor task, that is used in post-stroke rehabilitation trainings. When performing repetitive tasks with motor impaired patients, large performance variations can be observed not only over the course of a single session, but also on a single-trial level [22, 23]. Training over several sessions can reduce but not dissolve these task-related trial-to-trial variations [24]. The underlying mechanisms for the behavioral variability are not fully resolved yet. They might be explained by the stochastic character of the nervous system, which arises from noise in brain networks [25] or a specific regulation of variability to support motor learning [26]. While most BCI-based post-stroke motor rehabiliation trainings focus on the direct decoding of movement intentions [11], the ongoing brain state which might be informative about upcoming trial-to-trial performance fluctuations has not been taken into account so far.

1.1 RESEARCH CONTRIBUTIONS

Contributions

With this interdisciplinary thesis, brain signals prior to and during a repetitive motor task—which is utilized in post-stroke motor rehabiliation trainings—will be studied extensively. As trial-to-trial performance variations are expected in this context, it will be investigated if multivariate machine learning methods allow identifying robust individual neural features (markers) to explain and even temporally predict trial-by-trial motor performance variations. Specifically, neural oscillations will be targeted as a rich source of information [27]. Datasets on normally aged controls and stroke patients were collected, to validate whether individual performance predictors can be identified in both groups.

In patient scenarios current brain state decoding methods are stretched to their limits. Here, predominant data regimes are characterized by a high dimensionality, scarce training data and label noise conditions. Thus, it will be verified if a state-of-the-art machine learning approach can be robustified by deploying different regularization strategies to reduce the algorithms' susceptibility to overfitting the training data. Pushing forward the frontiers of model introspection, a novel data-driven framework by mining the oscillatory brain activity dynamics will be provided to identify

reliable and functionally relevant neural features. Here, the exploitation of within-trial event structure is foreseen as a key ingredient.

In a final step, it will be investigated if the robustified decoding methods can be directly transferred into a novel brain state-dependent interaction protocol exemplified for the previously studied hand motor task in post-stroke rehabilitation. For the first time, it will be verified whether single-trial motor performance can be influenced according to the individual pre-trial brain state estimates. Overall, this thesis contributes novel machine learning tools which help to increase the individual efficiency of closed-loop interaction protocols, for instance in rehabilitation scenarios.

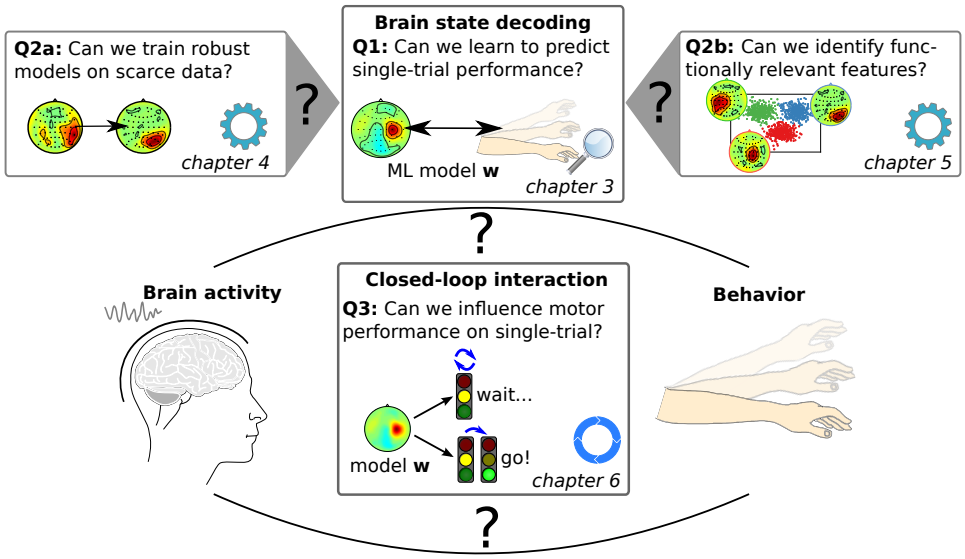


Figure 1.1: **Schematic thesis structure.** The integration of all posed research questions into a general context is schematically shown. Each research question is related to a specific chapter.

As summarized in Fig. 1.1, the research contributions of this thesis can be translated into three overall research questions which will be successively followed throughout this thesis:

Overall tackled research questions

- **Q1**: Can data-driven brain state decoding methods reliably predict single-trial motor performance from pre-trial non-invasive oscillatory brain activity?
- **Q2a**: How can such decoding approaches be robustified under scarce training data and label noise conditions?
- **Q2b**: Can we build data-driven decoding methods that focus on identifying reliable and functionally relevant features?
- **Q3**: Can upcoming single-trial motor performance be influenced in closed-loop interaction based on the ongoing brain state estimates?

1.2 OUTLINE

Hereafter, a short outline on the structure of this thesis is provided. The major contributions were previously published in peer-reviewed journals and conference proceedings. For each chapter, the corresponding publications are explicitly stated in the following.

Chapter 2 gives a brief summary on background knowledge about assessing human brain activity followed by a wrap-up of basic oscillatory brain signal analysis. Specifically, state-of-the-art spatial filtering approaches are shortly reviewed. In addition, the commonly applied BCI data processing pipeline is described. Furthermore, a short review on the most prominent BCI application fields is provided.

Chapter 3 presents an offline analysis framework to identify robust individual oscillatory features that allow to partially explain and temporally predict motor performance variability. For this purpose, a state-of-the-art supervised machine learning approach is utilized to regress a known continuous trial-wise variable from multivariate brain activity recordings. The proposed framework is established for a hand force task which was designed for post-stroke motor rehabilitation training [28–31].

State-of-the-art algorithms for continuous brain state decoding reach their limits, especially in patient scenarios. As features need to be optimized individually, substantial training data is required. To improve the performance with scarce training data, three types of regularization techniques are proposed and characterized for the investigated brain state decoding approach in chapter 4. For the evaluation, a novel simulation framework as well as two real-world datasets were considered [32, 33].

Current brain state decoding models mostly neglect contextual information of underlying experimental paradigms. Moreover, trained models are extremely sensitive to changes in training data or involved hyperparameters. This leads to highly variable solutions and impedes the selection of a proper model for closed-loop interaction. Fostering component introspection, chapter 5 proposes an unsupervised ML approach to identify reliable and functionally relevant oscillatory features by mining task-specific brain signatures [34].

Based on the identified objective neural markers for the decoding of task-specific suitable brain states, their final validation is targeted in chapter 6. The developed algorithmic advances of chapters 4 and 5 will be utilized in a novel brain state-dependent closed-loop training evaluated with chronic stroke patients. The studied hand motor training is complemented with a novel single-trial strategy exploiting the ongoing brain state in order to influence single-trial motor performance.

The thesis will be concluded in chapter 7 by a final summary of the overall findings, their relevance for other application scenarios and possible future directions.

1.3 PUBLICATIONS

The major content of this thesis has already been published in peer-reviewed scientific journals and conference articles. In the following, an overview of all scientific publications is given in reverse chronological order for the period of my PhD work. It is split in two main categories, namely first author contributions and further involvement to scientific projects of collaborators and colleagues.

1.3.1 *Main contributions*

JOURNAL ARTICLES

- **A. Meinel**, H. Kolkhorst, M. Tangermann, *Mining Within-Trial Oscillatory Brain Dynamics to Address the Variability of Optimized Spatial Filters*, IEEE Transactions on Neural Systems & Rehabilitation Engineering, Vol. 27, No. 3, pp. 378–388, (2019) doi: 10.1109/TNSRE.2019.2894914
- **A. Meinel**, J.S. Castaño-Candamil, B. Blankertz, F. Lotte, M. Tangermann, *Characterizing Regularization Techniques for Spatial Filter Optimization in Oscillatory EEG Regression Problems*, Springer Neuroinformatics, Vol. 17, No. 2, pp. 235–251,(2019) doi: 10.1007/s12021-018-9396-7
- **A. Meinel**, J.S. Castaño-Candamil, J. Reis, M. Tangermann, *Pre-Trial EEG-Based Single-Trial Motor Performance Prediction to Enhance Neuroergonomics for a Hand Force Task*, Frontiers in Human Neuroscience, Vol. 10, (2016) doi: 10.3389/fnhum.2016.00170

PEER-REVIEWED CONFERENCE ARTICLES

- **A. Meinel**, F. Lotte, M. Tangermann, *Tikhonov Regularization Enhances EEG-Based Spatial Filtering For Single-Trial Regression*, Proceedings of the 7th Graz Brain-Computer Interface Conference 2017, pp. 308-313 (2017)
- **A. Meinel**, J.S. Castaño-Candamil, S. Dähne, J. Reis, M. Tangermann, *EEG Band Power Predicts Single-Trial Reaction Time in a Hand Motor Task*, Proc. IEEE Conference on Neural Engineering (NER), pp. 182-185 (2015)

ABSTRACTS

- **A. Meinel**, T. Koller, M. Tangermann, *Time-Frequency Sensitivity Characterization of Single-Trial Oscillatory EEG Components*, The First Biannual Neuroadaptive Technology Conference, pp. 36-37 (2017)
- **A. Meinel**, K. Eggensperger, M. Tangermann, F. Hutter, *Hyperparameter Optimization for Machine Learning Problems in BCI*, Proc. Sixth Int. BCI Meeting, p. 184 (2016)

- **A. Meinel**, E. M. Schlichtmann, T. Koller, J. Reis, M. Tangermann, *Predicting Single-Trial Motor Performance from Oscillatory EEG in Chronic Stroke Patients*, Proc. Sixth Int. BCI Meeting, p. 140 (2016)

1.3.2 *Additional contributions*

- J.S. Castaño-Candamil, **A. Meinel**, M. Tangermann, *Post-Hoc Labeling of Arbitrary EEG Recordings for Data-Efficient Evaluation of Neural Decoding Methods*, Proc. Seventh Int. BCI Meeting, pp. 58-59 (2018)
- J.S. Castaño-Candamil, **A. Meinel**, M. Tangermann, *Post-Hoc Labeling of Arbitrary EEG Recordings for Data-Efficient Evaluation of Neural Decoding Methods*, arXiv e-prints (2017)
- M. Tangermann, **A. Meinel**, *Informative Oscillatory EEG Components and their Persistence in Time and Frequency*, Neurotechnix Vol. 1, CogNeuroEng pp. 17-21 (2017)
- J. Meyer, **A. Meinel**, T. Schreiner, B. Rasch, M. Tangermann, *Versuchspersonenunabhängige Single-Trial-Erkennung von langsamen Wellen im Schlaf-EEG*, 24. Jahrestagung der DGSM, Somnologie 20, pp. 75-76 (2016)
- M. Tangermann, J. Reis, **A. Meinel**, *Commonalities of Motor Performance Metrics are Revealed by Predictive Oscillatory EEG Components*, Proc. 3rd Int. Congress on Neurotechnology, Electronics and Informatics (Neurotechnix), pp. 32-38 (2015)
- J.S. Castaño-Candamil, **A. Meinel**, Sven Dähne, M. Tangermann, *Probing Meaningfulness of Oscillatory EEG Components with Bootstrapping, Label Noise and Reduced Training Sets*, Proc. 37th Int. Conf. of the IEEE Eng. in Medicine and Biology Soc. (EMBC) pp. 5159- 5162 (2015)
- J.S. Castaño-Candamil, **A. Meinel**, J. Reis, M. Tangermann, *Correlates to Influence User Performance in a Hand Motor Rehabilitation Task*, Clinical Neurophysiology, 126 pp. 166- 167 (2015)

2 FUNDAMENTALS

Hereafter, I provide a short introduction on basic methods and findings that are commonly shared by all chapters and essential to get access to the interdisciplinary topic. In each chapter, I will specifically provide the methodological background to integrate the corresponding chapter into the current state-of-the-art.

2.1 MEASUREMENT OF HUMAN NEURONAL ACTIVITY

To gain an understanding about the basic mechanisms of the human brain, it is indispensable to record activity of the nervous system. In the last century, a variety of functional brain imaging methods have been developed which cover a broad range of scales both in terms of spatial and temporal resolution [35]. On the microscopic level, there are techniques to capture spiking activity of single neurons or even single synapses. On the macroscopic level, activity of large neuronal populations can be recorded by the electroencephalography (EEG). This non-invasive method —first described by [36]—allows measuring changes in electrical scalp potentials by placing electrodes on the human scalp. As this thesis builds upon novel ML tools for single-trial EEG analysis, we will shortly review the electrophysiological basis of this method which can be found, among others, in Baillet, Mosher, and Leahy [37], Kandel et al. [38], Nunez and Srinivasan [39], and Hallez et al. [40].

2.1.1 *Electrophysiology of the EEG*

The human brain consists of about 10^{10} nerve cells or neurons which are strongly interconnected among each other. The junctions of two neurons are called synapses. Neurons share a common anatomical structure as they consist of a cell body (soma), the branching dendrites to receive input from other nerve cells and the axon for transmitting information to other cells.

Single neurons are electrically excitable as their cell membrane is composed of ionic pump proteins which actively regulate the ion transport across the membrane. This results in a concentration gradient of different ions and thereby a total charge difference between the intra- and extracellular membrane translates into the resting potential. This concentration gradient—and thus its corresponding resting potential—is actively maintained by the energy-consuming ionic pumps.

Neurons are capable to receive, process and transmit information by a change of the membrane resting potential. The process behind it is the following: when taking a closer look at a synapse, an active pre-synaptic neuron can release neurotransmitters which enter the dendritic tree of

a post-synaptic neuron. This synaptic input causes an ionic current flow across the membrane which results in a post-synaptic potential (PSP). A single post-synaptic neuron receives synaptic input from many other pre-synaptic nerve cells. These various PSPs are transmitted along the dendrites and become integrated at the soma. If the overall depolarization of the intracellular space is large enough, then voltage-sensitive ion channels suddenly open up. This causes an ionic flow—called primary current—and thus delivers an action potential which travels along the axon to other neurons.

According to the principle of charge conservation, there is also a secondary current through the extracellular space. As each current generates an electric field according to Ohm's law, these secondary currents are reported as the main generators of measurable scalp potentials that can be recorded with EEG. Two major factors influence the resulting measurable potentials: First, the spatial orientation of the cortical neurons are crucial. EEG signals are dominated by the electric fields of large, parallel oriented pyramidal cell assemblies orthogonal to the cortical surface. Second, synchronous activation of these cortical networks is crucial as both factors facilitate a superposition of the resulting electric fields. EEG recordings reveal typical amplitudes in the order of a few microvolts. Such amplitudes require synchrony along cortical networks over at least $10\,cm^2$ [41] and thereby providing a spatial resolution estimate of the EEG.

EEG signals are of interest in frequency ranges considerably below 1 kHz. Fortunately, in this range the physics of EEG allows neglecting any electromagnetic propagation effect. In other words, the mixture of various neural source activations translate without temporal delay into a measurable scalp potential. This effect is called volume conduction as the anisotropic conductive properties of the head's different layers, especially the cerebrospinal fluid and the skull, lead to a blurred scalp potential distribution.

EEG activity is typically measured by multiple electrodes that are placed at approximately equidistant locations across the whole scalp. For reproducibility, the placement follows a fixed scheme, such as the common 10-20 system. In order to assess a potential for each EEG electrode, the signal of a single EEG channel is obtained by referencing the corresponding electrode signal against a reference, such as a single physical electrode.

When recording EEG signals in practice, neuronal signal sources are superimposed by artifactual contributions. These are caused by non-neural physiological sources, e.g., ocular, muscular or cardiac activity and by non-physiological origins such as hardware related artifacts. This needs to be taken into account when analyzing such data [42].

2.1.2 *Oscillatory brain activity*

The spectral power density of the EEG— as well as signals obtained by invasive techniques such as local field potentials (LFPs) or electrocorticography (ECoG)—is inversely proportional to the frequency f. The characteristic

$1/f$ decrease in the power spectrum is generally superimposed by one or multiple narrow frequency-specific peaks as shown for an exemplary dataset in Fig. 2.1 (A). These peaks reflect oscillatory processes with increased rhythmic activity in subject-specific frequency bands such as the *alpha* band which can roughly be found in the range from $8-13$ Hz and the *beta* band in the range of $14-30$ Hz. Moreover, other neurophysiologically relevant frequency bands include the delta ($1-3$ Hz), the theta ($4-7$ Hz) and the gamma frequency band (30 to more than 100 Hz). These different frequency domains are linked to functional brain networks which are relevant for cognitive and physiological functions [43, 44]. A more detailed review on the origin and properties of oscillatory processes can be found in [27]. When measuring macroscopic oscillatory brain activity specifically at electrode locations in the immediate proximity of the sensorimotor cortex, such activity is commonly referred to as sensorimotor rythm (SMR) or μ rhythm.

The modulation of oscillatory activity was found to play a major role in various cognitive processes such as, among others, memory, attention or visual perception [45–48]. A prominent example is given by the modulation of alpha power by the visual system as it increases the power over parieto-occipital channels when a subject closes the eyes.

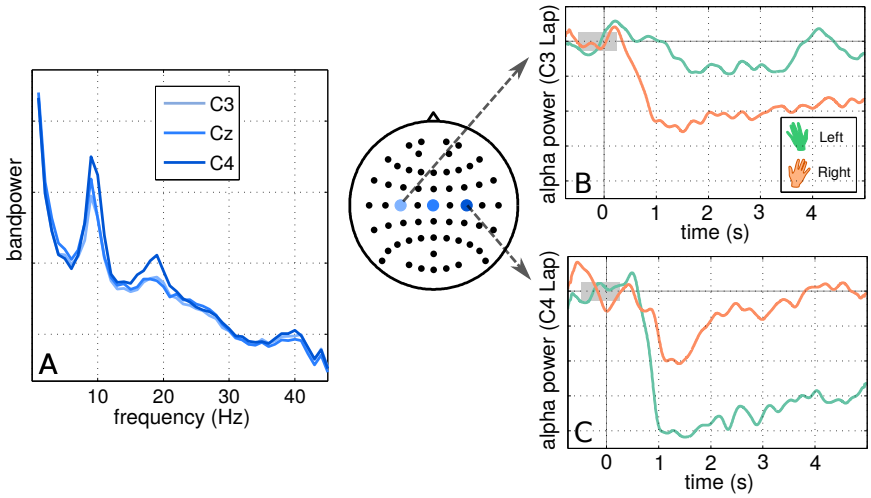


Figure 2.1: **Basic analysis of oscillatory brain activity.** Based on data of a single session, two key phenomena are shown: (1) The frequency spectra for three single EEG channels are shown in (A). (2) Event-related power modulation (here: ERD) for two contrasting conditions of left- and right-hand motor imagery. In (B) and (C), the power time course of Laplace filtered channels $C3$ and $C4$ are given. The time $t = 0$ refers to the onset of movement imagination.

A state-of-the-art approach to analyze oscillatory activity from EEG recordings is to verify the presence of time-locked, frequency-specific envelope modulations. They can be triggered by an internal or external event such as a simple visual cue. The induced power modulation effect is known as event-related (de-)synchronization (ERD/ERS) [49–51]. Such an effect

can be observed when analyzing the class-wise envelope dynamics for a motor imagery (MI) dataset. Instructed by, e.g., a visual cue, a subject is asked to imagine either left or right hand movements. Fig. 2.1 (B) and (C) report the alpha power time course for two Laplace filtered channels $C3$ and $C4$. Laplace filters are one of the simplest methods to enhance the typically low signal-to-noise (SNR) of EEG recordings by taking neighboring channels into account [52]. Typically, the channels $C3$ and $C4$ are directly located over the left and right primary motor cortex. As usually found for MI data [53], there is a SMR decrease less than 1 s after the movement onset, hence an ERD effect on the contralateral hemisphere. For the presented data, there is also a less strong ERD on the ipsilateral side observable.

Several approaches to extract and quantify ERD/ERS effects along the spatial, temporal and spectral domain have been proposed [54–57] and thus enable to identify involved cortical regions evoking such power changes [58, 59]. Studies on ERD/ERS effects are mostly implemented with simple tasks. These typically result in a simple, cue-locked modulation of the band power [50] such as the reported ERD upon movement onset in Fig. 2.1. More complex tasks with a richer temporal structure and a possibly richer resulting ERD/ERS structure have been studied far less frequently.

However, the extraction of distinct oscillatory signatures from Laplace filtered channels (as in Fig. 2.1) or even based on only single EEG channels is not always straightforward. In general, the high-dimensional EEG or similarly magnetoencephalography (MEG) recordings reveal an intrinsic low SNR ratio [60]. In addition, strong inter-individual differences in the time-frequency and spatial characteristics of SMR activity challenge the automation of such oscillatory analyses. To tackle these challenges, sophisticated signal processing and machine learning approaches come into play which will be introduced in Sec. 2.3.2.

Interestingly, most subjects can learn the ability to voluntarily modulate their SMR activity, e.g., by imagined movements. This capability supplies a direct neural pathway for controlling SMR-based BCI systems. Their general structure and functionality is explained hereafter.

2.2 BRAIN-COMPUTER INTERFACES

2.2.1 *Closed-loop interaction*

The commonly established closed-loop processing pipeline of an *online* BCI system is depicted in Fig. 2.2 [61]. In general, such systems collect, analyze and finally translate brain signal recordings into output commands in a real time. More precisely, five major steps are realized by an online BCI system: (1) Brain signals are continuously recorded by either an invasive or non-invasive imaging technique. In this thesis, we will restrict the view on non-invasive EEG-based BCIs. (2) Data preprocessing consists of steps such as removing artifactual signal contributions or temporal filtering of the data to a specific frequency band of interest. A proper preprocessing typically ensures that outliers are removed from the data. (3) Feature

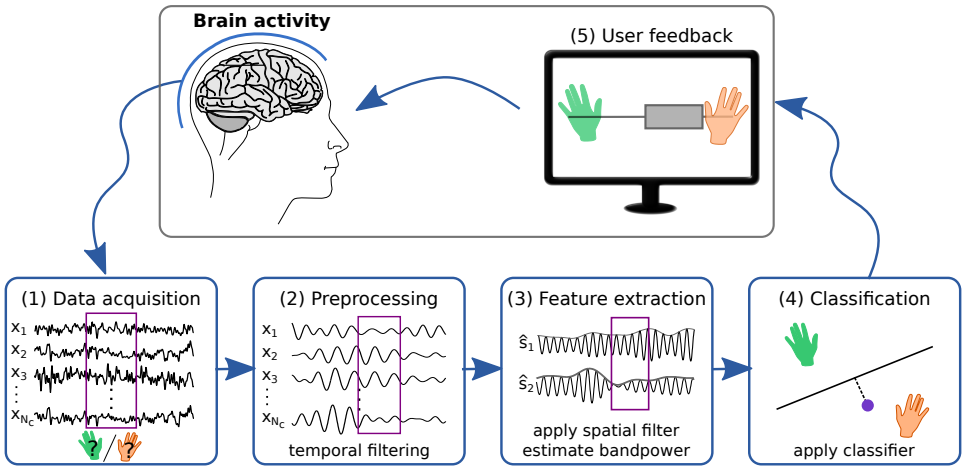


Figure 2.2: **Main processing steps for online BCI systems.** The scheme is exemplified for a motor task classification pipeline based on modulated oscillatory brain rhythms which are typically used as features.

extraction is performed by estimating brain signal components of interest, e.g., by applying a pre-trained model to extract task-related information. (4) A pre-trained classification or regression model is evaluated on the extracted features to decode the user's intention or brain state and finally feed the BCIs control signal into a feedback application. (5) In a final step, the loop is closed by providing feedback to the user via a common modality, such as the visual, auditory or tactile pathway or by utilizing the decoded information to trigger an additional device, for instance a robot, an orthosis or an electrical stimulation. The steps (3) and (4) require trained ML models for the decoding of relevant information from ongoing brain activity recordings. In many scenarios, supervised ML models are utilized such that labeled training data is required. The necessary training of the ML models—also known as *calibration*—is typically performed on sufficient data of a previous session or the initial phase of an ongoing session. Reducing the time for collecting a sufficient amount of calibration data is one major challenge in the field of BCI research and is a key factor to facilitate the use of BCI systems outside the laboratories [62].

2.2.2 *Clinical application fields*

So far, BCIs for clinical applications are predominantly studied along two lines of research which will be shortly introduced hereafter.

BCIs for communication and control

Initially, the majority of BCI research focused on the development of assisstive technology to restore communication and control capacities for severely impaired patients with chronic neuromuscular disorders, such as spinal cord injury, amyotrophic lateral scelorisis, among others. Such disorders damage the neural pathways for muscle control or directly impair

the muscles themselves. In the absence of almost any voluntary muscular control (including eye-movements), a BCI might be capable to restore function by decoding information from still intact brain areas and utilize it as a channel for communication and control. Thereby, the efferent pathways of the brain are bypassed [7, 63, 64]. Such applications are typically based on either stimulus-induced paradigms such as a *P*300 speller [65] or self-paced paradigms. The later one is typically realized by MI paradigms in which the class-specific SMR modulation (see Fig. 2.1) can be exploited to decode the user's intention [66].

BCI-based post-stroke (motor) rehabilitation

In the past decade, a growing number of BCI applications to support rehabilitation scenarios after stroke can be found. In general, such systems strive to close afferent and efferent neural pathways that are interrupted by the stroke [10, 12, 67]. This concept was not exclusively applied to motor rehabilitation scenarios but also for attention- or language-related deficits [68–70].

Restricting the view on BCIs for post-stroke motor rehabilitation, these systems conceptually aim to trigger functional and structural re-organization (neuroplasiticity) by reinforcing brain states that are beneficial for motor recovery. Most studies focus on the direct decoding of movement intention or motor imagery and thereby close the loop between efferent and afferent pathways. The specific way how individual sensory feedback is provided comprises a broad spectrum of applications. This can be done by functional electrical stimulation [71], the triggering of an external device such as a robotic arm [72] or an orthosis [15, 73] or by providing embodied feedback in virtual reality setups [74].

So far, research on BCI-supported motor rehabilitation is in a very early state, such that the efficiency of such training protocols including a control group has currently been reported in very few studies only [17].

2.3 SINGLE-TRIAL BRAIN STATE DECODING METHODS

As illustrated in Fig. 2.2, BCI systems typically require trained models which allow the decoding user intention or the extraction of individual informative brain states from ongoing multichannel brain signal recordings [61]. Often, these signals come with a high dimensionality and are typically superimposed of both task-related and unrelated neural contributions, given that a sufficient preprocessing was performed by removing outliers and artifactual signals components, among others. The still resulting low intrinsic SNR ratio of such signals impedes the extraction of hidden, informative neural sources. Moreover, brain signals are non-stationary, as they typically reveal a strong trial-to-trial variability within and across subjects.

To remedy the situation, algorithmic approaches that fuse information from all available channels —commonly referred to as *multivariate*

methods— from the fields of machine learning and signal processing have proven beneficial as they can provide access to different neural processes under improved SNR ratios [19]. In the following, multivariate brain signal analysis is shortly revisited by introducing the general mathematical concept of forward and backward modeling of macroscopic brain activity measurements [75, 76].

Hereafter, multivariate EEG recordings acquired from a total number of sensors or channels N_c at a time point t will be denoted by a vector $\mathbf{x}(t) = [x_1(t), ..., x_{N_c}(t)]^\top \in \mathbb{R}^{N_c}$. Collected data, e.g. over the course of an experimental session, with overall N_{sam} samples are concisely notated by the multivariate data matrix $\mathbf{X} = [\mathbf{x}(1), ..., \mathbf{x}(N_{sam})] \in \mathbb{R}^{N_c \times N_{sam}}$.

2.3.1 *Linear forward model*

In the following, we assume that observed multivariate data $\mathbf{x}(t)$ can be described as a mixture of unknown variables—named *components* or *factors*—which reveal a distinct spatio-temporal signature. Each component can thus be regarded as a hidden functional process, e.g., a stimuli-evoked brain response that is segregated from other neural sources that contribute to the finally observable data $\mathbf{x}(t)$. Moreover, the temporal activity at time t of the k^{th} component is captured by the scalar variable $s_k(t)$. The projection of a component's activity $s_k(t)$ to the individual scalp channels is described by the spatial activation pattern $\mathbf{a}_k \in \mathbb{R}^{N_c}$ which holds the strength and polarity of the projection. As such, an activation pattern $\mathbf{a}$ provides access to a neurophysiological interpretation, as it spatially describes to which scalp channels the source component is contributing.

The *linear forward/generative model* describes the linear superposition of $K \geq 1$ latent variables which are summarized by the vector $\mathbf{s}(t) = [s_1(t), ..., s_K(t)]^\top \in \mathbb{R}^K$ with $k \in \{1, ..., K\}$ to explain the observed multivariate data $\mathbf{x}(t)$ at time t. Now, each component is mapped to the sensor space by their corresponding spatial activation pattern [75]:

$$\mathbf{x}(t) = \sum_{k=1}^{K} \mathbf{a}_k \cdot s_k(t) + \epsilon(t) = \mathbf{A}\mathbf{s}(t) + \epsilon(t) \tag{2.1}$$

with $\epsilon \in \mathbb{R}^{N_c}$ referring to a noise term which captures the remaining signal part that is not yet explained by the K source components. The matrix $\mathbf{A} \in \mathbb{R}^{N_c \times K}$ contains all spatial patterns of corresponding latent sources.

2.3.2 *Backward model*

Now given the recorded data $\mathbf{x}(t)$ only, conversely the question emerges of how to find the corresponding source components. In the most general case, the two unknowns $\mathbf{A}$ and $\mathbf{s}(t)$ can be estimated jointly which is referred to as blind source separation (BSS). However, there is not a unique solution to this factorization. It even requires additional assumptions about the underlying temporal and spatial dynamics of the related brain sources of interest.

When aiming for a joint estimation of the spatial patterns $\mathbf{A}$ and the source activation time course $\mathbf{s}(t)$, this will involve a highly complex optimization problem. It can be substantially simplified by considering a linear *backward/inverse model*. Such approaches typically decompose the signal into $1 \leq K \leq N_c$ estimated source components $\hat{\mathbf{s}}(t)$ by a linear projection matrix $\mathbf{W} \in \mathbb{R}^{N_c \times K}$ such that [75, 76]:

$$\hat{\mathbf{s}}(t) = \mathbf{W}^\top \mathbf{x}(t) \tag{2.2}$$

Each column of the matrix $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_K]$ refers to a single spatial filter $\mathbf{w}_k$. Such a model is composed of a set of coefficients that determine the linear combination of the single recorded channels in order to estimate the time course of a corresponding source component $\hat{\mathbf{s}}_k(t)$. In the following, the subscript index will be discarded when referring to a single filter model.

2.3.3 *Neurophysiological model interpretation*

An important aspect about the estimation of any backward model is to gain an understanding of the underlying subspace component in the original sensor space. Interestingly, such linear methods allow for an interpretation of the subspace components as pointed out by Haufe et al. [76]. However, as the coefficients of a single spatial filter $\mathbf{w}$ are depending on both, the signal and noise structure in the data, a direct neurophysiological model interpretation is not possible for $\mathbf{w}$. For every backward model as declared in Eq. (2.2), there is an existing forward model according to Eq. (2.1). It can be shown that the corresponding spatial activation patterns $\mathbf{A}$ can be obtained from the spatial filter matrix $\mathbf{W}$ based on the averaged covariance matrix $\mathbf{\Sigma}_{avg}$:

$$\mathbf{A} = \mathbf{\Sigma}_{avg}\mathbf{W}(\mathbf{W}^\top \mathbf{\Sigma}_{avg}\mathbf{W})^{-1} \propto \mathbf{\Sigma}_{avg}\mathbf{W} \tag{2.3}$$

The last step holds if the estimated sources $\hat{\mathbf{s}}$ are uncorrelated—which can be stated for many backward model techniques— or for the case $K = 1$. Thus, the activation pattern can simply be estimated from the averaged covariance matrix of the data and the spatial filter estimate.

Given we can estimate a backward model $\mathbf{w}$ on a dataset, Eq. (2.3) allows us to estimate the corresponding forward model. Thus, we can limit further considerations on ways to formulate cost functions only depending on the filter $\mathbf{w}$.

2.3.4 *Data-driven spatial filtering algorithms*

The recovery of source components from observed data—as formulated in Eq. (2.2)—can also be tackled as a machine learning problem. In general, the task is then to learn a subspace decomposition of the data parametrized by the matrix $\mathbf{W}$. In current literature, there are various backward model approaches to estimate spatial filter coefficients in a data-driven manner and relate them to behavioral data [77]. However, their choice strongly depends on the final application scenario. A large number of state-of-the-art

algorithms for spatial filter estimation share a common mathematical formulation of the optimization function $f_{obj}(\mathbf{w})$ with an additional constraint $f_{cst}(\mathbf{w})$:

$$\max_{\mathbf{w}}/\min f_{obj}(\mathbf{w}) = \mathbf{w}^\top \mathbf{M}_1 \mathbf{w} \quad \text{s.t.} \quad f_{cst}(\mathbf{w}) = \mathbf{w}^\top \mathbf{M}_2 \mathbf{w} \stackrel{!}{=} q \tag{2.4}$$

with the matrices $\mathbf{M}_1$, $\mathbf{M}_1 \in \mathbb{R}^{N_c \times N_c}$ and a constant $q \in \mathbb{R}$. The exact definition of $\mathbf{M}_1$ and $\mathbf{M}_2$ is then specific to each algorithmic solution.

As Eq. (2.4) holds an equality constraint, the method of Lagrange multipliers can be utilized to transfer the formulation into a symmetric generalized eigenvalue problem (GEP) $\mathbf{M}_1 \mathbf{w} = \lambda \mathbf{M}_2 \mathbf{w}$ with eigenvalue λ and corresponding eigenvector $\mathbf{w}$. While the cost of a full eigenvector decomposition is of order $\mathcal{O}(N_c^3)$, efficient algorithms for solving GEPs are implemented in standard linear algebra toolboxes [78]. Solving a GEP typically returns a the full matrix $\mathbf{W}$ of N_c spatial filters. Hereafter, we assume a sorting of the K filters with $k \in \{1, .., K\}$ indexing the rank which can be determined in descending order of the eigenvalues.

In a general sense, spatial filtering algorithms can be subdivided into two main categories. *Unsupervised* methods solely consider statistical properties of the data in absence of any label information, while *supervised* techniques additionally exploit label information to explicitly direct the parameter optimization. The most relevant methods for this thesis will shortly be reviewed hereafter [79, 80].

2.3.4.1 *Principal component analysis*

The principal component analysis (PCA) is an unsupervised approach that optimizes for subspace components that explain the largest variations in the data [81]. The variance of a single source component can be expressed as $\mathrm{Var}(\mathbf{w}^\top \mathbf{X}) = \mathbf{w}^\top \Sigma \mathbf{w}$ with the covariance matrix $\mathbf{\Sigma} = (N_{sam} - 1)^{-1} \mathbf{X}^\top \mathbf{X}$. Thus, maximizing the variance under the constraint $||\mathbf{w}||^2 = \mathbf{w}^\top \mathbf{w} \stackrel{!}{=} 1$ of orthogonality leads to the following optimization problem:

$$\max_{\mathbf{w}} f_{obj}(\mathbf{w}) = \mathbf{w}^\top \mathbf{\Sigma} \mathbf{w} \quad \text{s.t.} \quad f_{cst}(\mathbf{w}) = \mathbf{w}^\top \mathbf{I} \mathbf{w} \stackrel{!}{=} 1 \tag{2.5}$$

with the identity matrix $\mathbf{I} \in \mathbb{R}^{N_c \times N_c}$. The direct comparison with the general constraint optimization formulation in Eq. (2.4) yields that the matrix $M_1 = \mathbf{\Sigma}$ and $M_2 = \mathbf{I}$. Due to the identity matrix, the formulation can be translated into a eigenvalue decomposition $\mathbf{\Sigma w} = \lambda \mathbf{w}$. As PCA components are orthogonal, they are also not correlated among each other. It can also be derived that the explained variance of a component is directly given by the corresponding eigenvalue $\lambda = \mathrm{Var}(\mathbf{w}^\top \mathbf{x}(t))$. Given a full PCA decomposition $\mathbf{W}$ with K PCA components, the mentioned eigenvalue property is of specific use to select a subset of $J \leq K$ components that correspond to the overall fraction of explained variance by $\sum_{j=1}^{J} \lambda_j / \sum_{k=1}^{K} \lambda_k$.

PCA is the most prominent spatial filtering approach and is utilized in various domains beyond the field of BCI for data visualization or feature extraction. In principle, unsupervised spatial filtering algorithms such

as PCA are commonly applied to reduce the dimensionality of usually high-dimensional multivariate brain signal recordings [82], while in other neuroimaging scenarios PCA is utilized to identify and project out artifactual contributions of observed data [83].

2.3.4.2 *Independent component analysis*

Another example of an unsupervised factorization approach to determine a set of spatial filters $\mathbf{W}$ is given by the independent component analysis (ICA). Instead of computing orthogonal components as accomplished in PCA, the ICA approach follows a different optimization strategy by identifying latent source components that are statistically independent among each other.

Formally, statistical independence of K random variables $\hat{\mathcal{S}}_i$ with $i \in \{1, ..., K\}$ can be assessed by the information-theoretic measure of mutual information $\mathcal{I}$. Under the assumptions of a linear mixing, we treat the source estimate $\hat{\mathcal{S}}$ as a set of K random variables which reveal the time series data of single source components according to the inverse model $\mathbf{W}$ introduced in Eq. (2.2). Given K source components, their mutual information $\mathcal{I}(\hat{\mathcal{S}})$ can be expressed as:

$$\mathcal{I}(\hat{\mathcal{S}}) = \sum_{i=1}^{K} \mathcal{H}(\hat{\mathcal{S}}_i) - \mathcal{H}(\hat{\mathcal{S}}) \geq 0 \tag{2.6}$$

with $\mathcal{H}(\hat{\mathcal{S}}_i)$ expressing the entropy of the variable $\hat{\mathcal{S}}_i$. It can be shown, that $\mathcal{I}(\hat{\mathcal{S}})$ becomes zero if and only if all K variables are statistically independent. Thus, striving for statistical independence, ICA is based upon minimizing the mutual information $\mathcal{I}(\hat{\mathcal{S}})$. One can show that minimizing $\mathcal{I}(\hat{\mathcal{S}})$ is equivalent to minimize the entropy $\mathcal{H}(\hat{\mathcal{S}}_i)$ for all single components. Given a fixed mean and variance of all random variables, the Gaussian distribution holds the maximal entropy. Thus, ICA finally favors K spatial filter estimates $\mathbf{w}$ that elicit the most non-gaussian source components $\hat{\mathcal{S}}_i$ [84, 85]. In consequence of the optimization based on mutual information, ICA is not transferable into the introduced GEP framework of Sec. 2.3.4.

When describing source components solely as random variables, the temporal information contained in brain signal recordings is completely discarded. However, there are ICA variants which takes the temporal source information into account [86].

ICA is applied in various scenarios for analyzing neuroimaging data [87, 88]. In the context of multivariate brain signal analysis, it is commonly used to identify artifactual components [89] or to extract informative neural components that reflect, e.g., task-related ERD/ERS effects [57, 59, 90].

2.3.4.3 *Prerequisites for supervised methods*

Based upon the course of events in an experimental paradigm, a multivariate recording $\mathbf{x}(t)$ can be translated into N_e single epochs with a corresponding data matrix $\mathbf{X}(e) \in \mathbb{R}^{N_c \times N_s}$ with N_s sample points per epoch.

By definition, a *supervised* machine learning model takes label information into account. As such, hereafter we assume to have access to an epoch-wise defined external target variable $z(e)$. This scalar variable z provides additional information about the underlying experimental paradigm and could represent, among others, stimulus characteristics, behavioral responses, external physiological information or cognitive measures.

Let $\mathbf{x}(t)$ and $\mathbf{X}(e)$ be bandpass-filtered to a frequency band of interest. Given a single spatial filter $\mathbf{w}$, the epoch-wise bandpower $\Phi(e)$ of the corresponding source $\hat{s}$ can be approximated by its variance:

$$\Phi_{\hat{s}}(e) = \mathrm{Var}[\hat{s}(t)](e) = \mathrm{Var}[\mathbf{w}^\top \mathbf{x}(t)](e) = \mathbf{w}^\top \boldsymbol{\Sigma}(e)\, \mathbf{w} \tag{2.7}$$

where $\Sigma(e) = (N_s - 1)^{-1} \mathbf{X}(e)^\top \mathbf{X}(e)$ denotes the epoch-wise spatial covariance matrix. Correspondingly, the average spectral power across all N_e epochs can be denoted as:

$$\Phi_{avg} = \langle \mathbf{w}^\top \boldsymbol{\Sigma}(e) \mathbf{w} \rangle = \mathbf{w}^\top \langle \boldsymbol{\Sigma}(e) \rangle \mathbf{w} = \mathbf{w}^\top \boldsymbol{\Sigma}_{avg} \mathbf{w} \tag{2.8}$$

with $\langle \cdot \rangle$ referring to an average across epochs and $\boldsymbol{\Sigma}_{avg} = \langle \boldsymbol{\Sigma}(e) \rangle$ defining the averaged covariance matrix across all epochs.

2.3.4.4 *Common spatial patterns*

A widely utilized multivariate method for solving a single-trial EEG classification task is tackled by the common spatial pattern (CSP) algorithm. It exploits underlying ERD/ERS effects induced by two contrasting experimental conditions, as encountered in a motor imagery paradigm [91–93].

In CSP, the epoch-wise label information is defined as a binary variable $z(e) \in \{0, 1\}$ and thus targeting a two-class problem. Such a label can for instance refer to a left or right hand movement imagination. In a nutshell, CSP optimizes for spatial filter estimates which maximize the contrast in terms of epoch-wise variance between the two classes. As the epoch-wise variance of a bandpass-filtered signal approximates the bandpower, it thereby extracts a subspace which maximally contrasts ERD/ERS effects among the two classes. Given the class-wise averaged covariance matrices $\boldsymbol{\Sigma}_c = \langle \boldsymbol{\Sigma}(e) \rangle_{z(e) \in c}$ with classes $c \in \{0, 1\}$, the following objective function can be stated:

$$\max_{\mathbf{w}} f_{obj}(\mathbf{w}) = \mathbf{w}^\top (\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_0)\, \mathbf{w} \quad \text{s.t.} \quad f_{cst}(\mathbf{w}) = \mathbf{w}^\top \boldsymbol{\Sigma}_{avg}\, \mathbf{w} \stackrel{!}{=} 1 \tag{2.9}$$

with $\boldsymbol{\Sigma}_{avg}$ referring to the overall averaged covariance matrix regardless of the class label. Comparing Eq. (2.9) with the general formulation in Eq. (2.4), we can identify $\mathbf{M}_1 = (\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_0)$ and $\mathbf{M}_2 = \boldsymbol{\Sigma}_{avg}$. Thereby, the CSP optimization can be translated into a GEP and thus delivers a closed-form solution. When estimating the full spatial filter matrix $\mathbf{W}$ by solving the GEP, typically the filters are sorted in decreasing order of the eigenvalues. According to the objective function in Eq. (2.9), the spatial filters $\mathbf{w}$ corresponding to the largest and smallest eigenvalue are the subspace

projections with maximize the variance ratio among the two classes. Thus, in a final BCI classification scenario one typically chooses a set of filters from both ends of the eigenvalue spectrum.

In summary, the CSP algorithm is commonly applied in oscillatory signal analysis when contrasting two experimental conditions. Finally, pre-trained CSP subspace components are utilized as features when running realtime BCI control applications [66] or when decoding motor system activations to support post-stroke motor rehabilitation [72].

2.3.4.5 *Source power comodulation (SPoC)*

The source power comodulation (SPoC) approach was recently proposed by Dähne et al. [94] for single-trial bandpower regression of multivariate oscillatory brain signals. The epoch-wise target variable $z(e)$ is now considered as a *continuous* variable. z is assumed to be standardized to zero mean — such that $\langle z(e) \rangle = 0$—and unit variance.

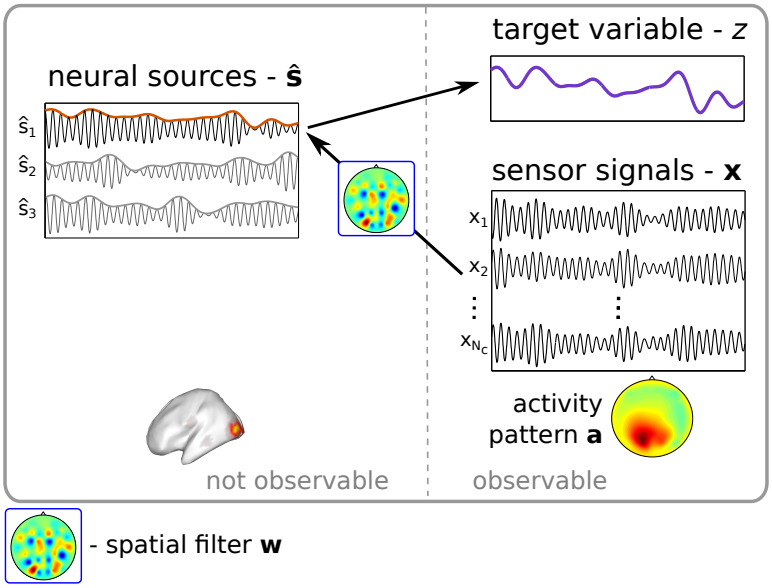


Figure 2.3: **Scheme of the SPoC optimization principle.** The method solves a single-trial regression problem based on multivariate observable sensor data $\mathbf{x}(t)$ and a continuous variable z. It optimizes for a spatial filter $\mathbf{w}$ such that the corresponding estimated neural source $\hat{s}$ maximally co-modulates with a given epoch-wise defined univariate target variable z.

As sketched in Fig. 2.3, the central idea of the SPoC algorithm is to search for an optimal spatial filter $\mathbf{w}^*$ such that the epoch-wise power $\Phi_{\hat{s}}(e)$ of the resulting estimated source component $\hat{s}$ maximally co-modulates with

the given target variable $z(e)$. Formally, this translates into the following objective:

$$\begin{aligned} f_{obj}(\mathbf{w}) &= \mathrm{Cov}[\Phi_{\hat{s}}(e), z(e)] \\ &\overset{Def.}{=} \langle (\Phi_{\hat{s}}(e) - \Phi_{avg})(z(e) - \underbrace{\langle z(e) \rangle}_{=0}) \rangle \\ &\overset{\text{Eq. (2.7, 2.8)}}{=} \mathbf{w}^\top \langle z(e)\, \mathbf{\Sigma}(e) \rangle\, \mathbf{w} - \underbrace{\langle z(e) \rangle}_{=0}\, \mathbf{w}^\top \mathbf{\Sigma}_{avg}\, \mathbf{w} \\ &= \mathbf{w}^\top \mathbf{\Sigma}_z\, \mathbf{w} \end{aligned} \tag{2.10}$$

where $\mathbf{\Sigma}_z := \langle z(e)\, \Sigma(e) \rangle$ defines the label-weighted covariance matrix averaged across epochs.

The original SPoC formulation by [94] comprised two different optimization strategies, namely covariance or correlation. This thesis operates on SPoC$_\lambda$ which optimizes covariance and allows deriving closed-form solutions for the spatial filters. Thus, the term SPoC will refer to the SPoC$_\lambda$ algorithm hereafter.

As the covariance is directly affected by the scaling of its arguments, it requires a constraint upon possible solutions. This is tackled by the previously utilized filter norm constraint of the CSP algorithm $\Phi_{avg} = \mathbf{w}^\top \mathbf{\Sigma}_{avg} \mathbf{w} \overset{!}{=} 1$. This leads to final objective function with a norm constraint:

$$\max_{\mathbf{w}} f_{obj}(\mathbf{w}) = \mathbf{w}^\top \mathbf{\Sigma}_z\, \mathbf{w} \quad \text{s.t.} \quad f_{cst}(\mathbf{w}) = \mathbf{w}^\top \mathbf{\Sigma}_{avg}\, \mathbf{w} \overset{!}{=} 1 \tag{2.11}$$

A comparison with Eq. (2.4) finally yields that $M_1 = \mathbf{\Sigma}_z$ and $M_2 = \mathbf{\Sigma}_{avg}$. Thus, the SPoC formulation can be translated into the known GEP framework and thereby delivers a closed-form solution. Overall, the approach returns a full set $\{\mathbf{w}_k^*\}_{k=1,..,N_c}$ of N_c spatial filters with k indexing the rank, which is determined in descending order of the eigenvalues and thereby according to the covariance.

Interestingly, the CSP formulation given in Eq. (2.9) can be directly derived from the SPoC formulation in Eq. (2.11) by defining a binary target variable $z \in \{0, 1\}$ instead of a continuous one.

2.3.5 *Applicability of spatial filters for closed-loop applications*

Finally, a short statement about the applicability of spatial filtering methods in closed-loop applications is given, as this will be a major research direction of this thesis.

Three major advantages are provided by the introduced linear spatial filtering methods [53, 79]: First, the approaches enhance the intrinsically low SNR ratio of high-dimensional brain activity recordings such as M/EEG and provide a data-driven way to identify discriminative information. Second, these linear models directly provide neurophysiological interpretability, which supports the understanding about underlying brain mechanisms and favors their use for clinical applications. Third, the methods need to

be directly applicable at low computational effort in online BCI systems to allow for single-trial brain state interaction, which is one of the ultimate goals of this thesis. When using a spatial filter model in an online BCI system, a preceding bandpass filtering is usually required. However, this filtering step is time consuming as each M/EEG channel needs to be filtered individually. In case of utilizing infinite impulse response (IIR) filters, such as Butterworth filters, this operation is strictly linear. Thus, fortunately the two operations become interchangeable such that the spatial filter/s can be applied prior to bandpass filtering which reduces the computational effort in online BCI systems and thus promotes the online applicability of spatial filter methods in combination with IIR bandpass filtering [95].

Several methods can be used to estimate source components (as introduced in Sec. 2.3.2), among them source reconstruction techniques. Such methods may provide a high level of interpretability for the results directly in the source space, and may describe non-stationarities in the data and other complex dynamics [96]. However, there are three potential drawbacks of source reconstruction approaches [97]: First, the estimation of $\hat{\mathbf{s}}(t)$ with state-of-the-art algorithms usually creates a rather high computational burden [98]. Second, the methods require either a forward model $\mathbf{A}$ for each individual subject, which may not be available in most situations since it corresponds to the exact acquisition of a subject's brain anatomy. Alternatively, a standardized head model [99] can be considered at the expense of outcome precision. Third, source reconstruction problems are intrinsically ill-posed, because the solution is a priori not unique (more unknowns compared to number of recording sites) and might also be non-stable due to the sensitivity to small changes in the training data. Thus, the quality of an estimated source depends on additional assumptions, such as the density of sources or their location within the brain.

The estimation of informative source components can also be achieved with non-linear models, such as neural network approaches [100], which may come up with higher decoding accuracies. However, their degree of model interpretability is still rather limited and currently under further investigation [101].

Part II

RESEARCH CONTRIBUTIONS

Hereafter, the different research contributions of this PhD thesis are presented in four chapters. An overview on the tackled research question can be found in Sec. 1.1.

3 SINGLE-TRIAL MOTOR PERFORMANCE PREDICTION

This chapter closely follows the journal publication "Pre-Trial EEG-Based Single-Trial Motor Performance Prediction to Enhance Neuroergonomics for a Hand Force Task" by Meinel et al. [28]. For this collaborative work, I have taken the lead in designing the experiment, in collecting the two datasets, in performing the comprehensive data analysis, in visualizing the results and in writing the manuscript. For this thesis, the analysis was enriched by an additionally collected dataset on chronic stroke patients.

SUMMARY

A framework for building electrophysiological predictors of single-trial motor performance variations is presented, exemplified for a repetitive hand force task. Based on collected datasets of normally aged controls and chronic stroke patients, strong trial-by-trial performance variations for five clinically relevant metrics were found. In an online simulation, the supervised multivariate regression approach SPoC was applied on EEG data of a short time interval prior to the start of each trial. For 16 out of 25 subjects, SPoC revealed robust oscillatory EEG subspace components, whose bandpower activity are predictive for the performance of the upcoming trial. Since SPoC may overfit to non-informative subspaces, we propose to apply three selection criteria accounting for the meaningfulness of the features. Across all subjects, the obtained components were spread along the frequency spectrum and showed a variety of spatial activity patterns. In summary, we identified subject-specific predictors that explain up to 36 % of the performance fluctuations and may serve for closed-loop experimenting.

3.1 INTRODUCTION

3.1.1 *Motivation*

Motor training is utilized in rehabilitation scenarios to accelerate the re-gain of lost motor function after brain injury [102]. State-of-the-art rehabilitation concepts are based on repetitive training tasks with the aim to reach a functional gain [103–105]. Most prominent training paradigms comprise mirror training [106], constraint-induced movement therapy [107], simultaneous bilateral training [108], BCI-supported training [10] and robot-assisted techniques [109]. Recent rehabilitation approaches include the training of novel,

unfamiliar motor skills instead of training well-known habitual motor tasks, attempting to optimize functional cortical reorganization.

Repetitive paradigms allow for the assessment of motor performance on a very fine-granular time scale. The performance of each single trial can be monitored by metrics such as the length, speed or smoothness of the produced movement trajectory. The distributions and temporal characteristics of trial-wise motor performance variations have been studied by different groups [26, 110–112]. While practicing a motor task over several sessions enables a user for skill acquisition [113], trial-by-trial variability of motor performance is a prominent feature which does not fully vanish with training [23, 24, 114]. The underlying neural mechanisms of motor performance fluctuations on short time scales is subject of controversial discussion in literature and is not fully resolved yet [24, 25, 111].

In the present work, we aim towards closing this gap. Therefore, trial-wise performance fluctuations of a sequential visuo-motor task (SVIPT, [115–117]) are investigated while registering a user's brain activity by EEG. In SVIPT trials, the quality of a movement changes within seconds and from repetition to repetition.

Research question Q1

As stated in Sec. 1.1, our *hypothesis* is that subject-specific pre-trial brain signals can partially explain and temporally predict the trial-by-trial fluctuations of the upcoming motor performance. Given that such informative neural markers exist, then the SVIPT paradigm could be altered in order to meet the cognitive ergonomic requirements of each single user. Practically, the starting time point of the upcoming trial can be determined based on the information contained in this pre-trial neural marker. Ideally, such a neuroergonomic closed-loop gating strategy could provide control over the level of difficulty. This might allow to causally influence user performance and ultimately support SVIPT motor learning in the long run.

In this chapter, a simulated online analysis for the extraction of robust and meaningful EEG components is developed. Precisely we evaluate, if the information contained in selected components is able to partially explain the trial-by-trial variation of SVIPT performance in a predictive fashion, i.e. the pre-trial component is required to predict the outcome of the upcoming trial. Moreover, the characteristics of the best performance predictors are investigated in a group-level analysis.

3.1.2 *Related work*

Paradigms which include *brain state-dependent experimenting* require that an informative neural marker can be extracted robustly from brain signals [20, 21]. Given the high dimensionality and noisy characteristic of most types of brain signals, the extraction and decoding of such individual neural markers is a challenging task [118].

Screening literature on relevant neural markers of visual and motor performance, it is important to make a distinction between the use of single-trial decoding in contrast to the extraction of statistical differences, which may even be reported as group averages. Neural features which

correlate with the task performance on the grand average (GA) of a set of subjects have limited usefulness for closed-loop experimenting with a given individual. As inter-subject differences get lost during the averaging, GA features may have low predictive power when tested with data of a novel subject. Research in the field of brain-computer interfaces (BCI) has pushed forward methods for single-trial decoding of individual brain activity (mostly EEG signals) [8, 18]. Results from this field affirm that brain signals and informative features vary strongly between individuals [19]. To obtain optimal decoding results, BCI data processing pipelines strive to identify subject-specific informative features. Technically, these are gained either from a calibration recording prior to the online use of the BCI [119], or by transfer learning methods [120] which exploit features from pre-trained machine learning models of earlier sessions or previous users. Furthermore, attention needs to be paid to temporal dependencies: brain features may correlate with previous, simultaneous or may even be predictive for future behavior. Only the latter brain features can serve as a tool for brain state-dependent experimenting.

Correlates of perception performance

Statistical correlates of visual perception performance are reported by several groups. For stimuli near the perception threshold, the pre-stimulus occipital alpha bandpower of the EEG correlates with the detection performance [46], even on a single-trial basis using predictive features [121]. In addition to bandpower, the pre-stimulus alpha phase was reported to correlate with the detection performance [47]. Single-trial decoding methods were not applied in those auditory studies, but the reported correlates precede the perception, which may open the possibility for closed-loop experimenting. Based on the findings of Hanslmayr et al. [121] and Dijk et al. [46], there are two examples that set up an online experiment based on occipital alpha bandpower features. Tonin et al. [122] using EEG data and Horschig et al. [123], who employed MEG signals, both decoded covert visual attention in a closed-loop experiment by utilizing single-trial feedback on the detected attention shift. However, both groups did not fully close the loop, e.g., by manipulating the perception performance, which may have been possible by selecting suitable brain states for stimulation. Gonzalez Andino et al. [124] studied a cued reaction time task and identified that gamma band oscillatory activity observed in fronto-parietal regions prior to the stimulus onset correlates with reaction time. Similarly, Hoogenboom et al. [125] stated that the strength of visually induced gamma band activity is predictive for the detection of stimulus motion. Somatosensory stimuli of low-intensity, but above threshold were delivered and combined with a distracting masker stimulus by Schubert et al. [126]. Investigating perceived vs. missed stimuli in an offline analysis, pre-stimulus beta bandpower over the left frontal cortex was found predictive for the perception performance on the grand average, as well as mu and beta bandpower over the pericentral sensorimotor areas.

Hand kinematics decoding (not prediction)

In the motor domain, several groups have successfully decoded hand kinematics, using the center-out task [127] as the dominating experimental approach. In their own work, Jerbi et al. [128] provide a review over the decoding of hand movement parameters such as direction, position and velocity based on brain signals. ECoG signals were used by Pistohl et al. [129] to decode two-dimensional hand movement trajectories using an autoregressive filtering approach. More recently, the decoding of continuous grasp kinematics from invasive brain signals has been shown by Flint et al. [130].

Considering non-invasive techniques, Waldert et al. [131] have decoded (but not temporally predicted) the hand movement direction based on MEG and EEG. Neural correlates which encode the velocity of a movement have been investigated by [132]. The decoding of produced grip force based on a phase feature extracted from the beta range has been reported on data of three subjects by [133]. Zaepffel et al. [134] reported an increased centro-parietal beta power during the planning period of grasping movements, but it was not investigated, if decoding may work on the basis of single trials. Focusing on single-trial methods, Lew et al. [135] used slow cortical potentials of the EEG from fronto-parietal areas to predict self-paced movement directions a few hundred milliseconds prior to movement onset. Similarly, Hammon et al. [136] inspected predictive EEG features for planning target directions using a cue-based paradigm.

Motor performance prediction

In the field of BCI research, Maeder et al. [137] studied a motor imagery paradigm. The single trial decoding performance of left vs. right hand movement imagery tasks could be correlated to the level of pre-trial alpha bandpower over the sensorimotor cortices. Despite used offline, this neural marker would allow for a predictive intervention in a closed loop. In their statistical analysis, Yang et al. [138] identified frontal alpha and beta bandpower features which correlate with performance metrics of a reaching task. Proceeding to single-trial methods, Meyer et al. [139] reported on data of six subjects, who performed a hand positioning task. Their offline analysis revealed that the normalized time-to-target could be predicted based on pre-cue alpha-band activity of the EEG.

The state-of-the-art can be summed up as follows: In the perception domain, several studies have established single-trial performance prediction, partially even in closed-loop applications. The situation is different for the motor domain since only very few studies have investigated subject-specific motor performance prediction in single-trial upon a sufficiently large subject group. Closest to all of these requirements is the study by [139]. Our research hypothesis builds exactly upon this point. In the context of a hand force task, we propose a generalized *workflow* which identifies subject-specific *predictive* oscillatory EEG features evaluated on a single trial basis.

3.2 METHODS

3.2.1 *EEG-tracked hand force task*

In the context of hand motor skill learning, Reis et al. [115] introduced the Sequential Visual Isometric Pinch Task (SVIPT), which demands an isometric force control of thumb and index finger. Interestingly, training-induced improvement of the SVIPT generalizes well to other hand motor control tasks, even though pinch grasp activities are rarely displayed during natural behavioral patterns. Compared to the original SVIPT setup, brain activity was additionally recorded using EEG during a training session for posthoc offline analysis. As introduced in our earlier conference article [29], the EEG-tracked SVIPT setup is sketched in Fig. 3.1.

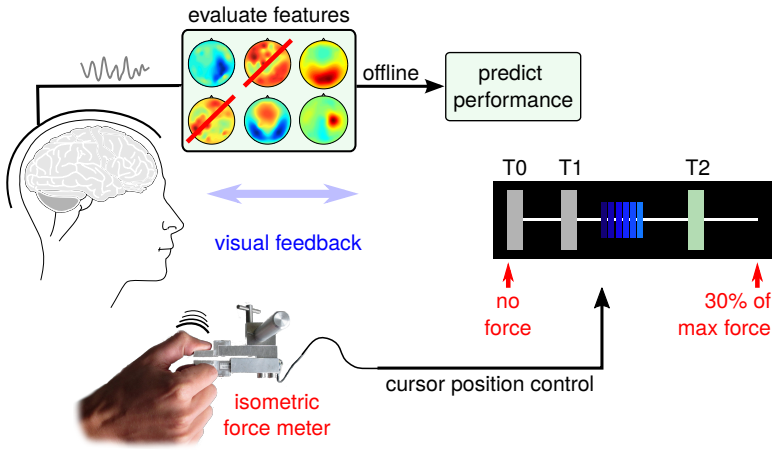


Figure 3.1: **Schematic setup of the offline EEG-tracked SVIPT.** The subject applies force to a sensor using a pinch grasp. The current force level is translated into a horizontal cursor position, while brain activity is recorded by EEG throughout the complete session.

Each SVIPT trial consists of three phases: a light blue (inactive) cursor appears on the leftmost edge of the $T0$ field (corresponding to zero force), while the user is touching the sensor only slightly with his hand. The appearance of a light blue cursor indicated the start of the *get-ready* phase, which corresponds to a waiting period with enhanced attention level. Its duration is varied uniformly between 2 and 3 s. The transduction of force into cursor movements is deactivated during the get-ready phase. Fixating the cursor, the user will observe a distinct color change of the cursor from light to dark blue. This *go-cue* indicates the beginning of the *running* phase, in which the cursor position can be controlled by applying force to the sensor. As force is transduced into horizontal cursor position, increasing force will move the dark blue cursor to the rightmost position at F_{limit}, which is pre-calibrated at session start to represent 30 % of the user's maximum force. The user has been instructed to navigate the cursor as quickly and accurately as possible by passing through a sequence of target fields ($T0$, $T1$ and $T2$). The target fields $T1$ and $T2$ were placed at 0.2 and

Table 3.1: **Overview on datasets of offline EEG-tracked SVIPT sessions.** Each dataset contains one offline session of each subject.

	dataset (D1a)	**dataset (D1b)**
Subject group information		
Subject group	norm. aged subjects	chronic stroke patients
No. of subjects (after preprocessing)	20 (18)	7 (7)
Age (years)	53±6	69±13
Gender (female/male)	8/12	3/4
Affected limb (left/right)	-/-	5/2
Experimental setup		
Recorded EEG channels N_{ch}	63	63
Trials per session N_e	400	240

0.8 of F_{limit}, respectively. Overshoots of the cursor had to be avoided. The current target field is visually indicated to the subject by a green shading (see Fig. 3.1), while the remaining ones are shaded in gray. Reaching a target field, a dwell time of 200 ms must be fulfilled in order to achieve a successful hit of this target field. Hit events are indicated visually by a switch of the target field (another field is shaded in green), or by the end of the trial. Trials were chosen randomly from two conditions, each with a specific required target field sequence ($T1$-$T0$-$T2$-$T0$ or $T2$-$T0$-$T1$-$T0$). A trial was finished by fulfilling the complete sequence – skipping a target was not allowed. The total trial duration including 1 s penalties for each overshoot were presented visually as an immediate performance feedback during the *pause* phase between trials.

3.2.2 *Recorded datasets*

A single session of the offline EEG-tracked SVIPT was performed with two different subject groups as summarized in Tab. 3.1.

Dataset (D1a) comprises 20 right-handed *normally aged* subjects. The group resembles the target group of first-stroke patients with respect to age and gender [140]. The term *normally aged* was chosen to indicate our selection criteria: the participants did not have any known neurological or psychological history and were probably healthy — even though we can not exclude the possibility that some participants had a history of unrecognized micro stroke events. In one session of about 3 to 4 hours, every participant controlled the cursor with their non-dominant left hand for 20 blocks of 20 trials each.

Dataset (D1b) comprises a set of chronic stroke patients with a first ischemic stroke more than 3 months before participation. All patients had a mild to moderate hemiparesis resulting in residual hand function which allowed to perform the foreseen hand force task. Here, the time per session was restricted to a maximum of 3 hours, such that every patient performed SVIPT for 12 blocks of 20 trials each.

The two offline studies were approved by the ethics committee of the University Medical Center Freiburg. Following the principles of the Declaration of Helsinki, written informed consent was given by subjects prior to participation.

3.2.3 *Single-trial motor performance metrics*

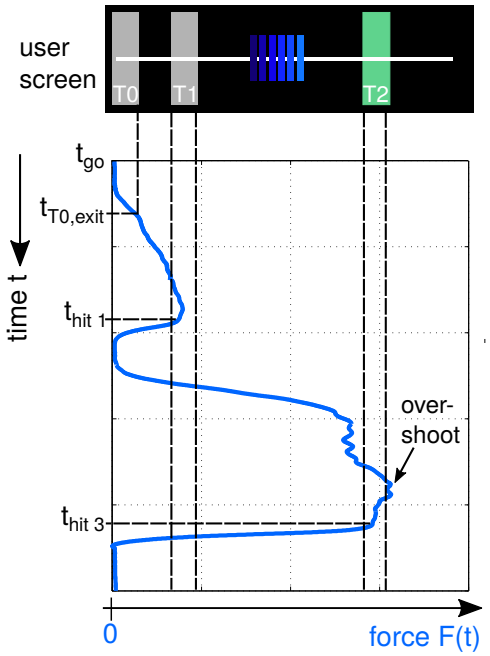


Figure 3.2: **Exemplary force profile $F(t)$ of a single SVIPT trial.** After the *go-cue*, the cursor is activated and can leave the target field $T0$ by applying force to the sensor. Different events are marked along the time axis. An overshoot event is highlighted, as the cursor has exceeded the target field $T2$. In this exemplary trial, the target field sequence was $T1$-$T0$-$T2$-$T0$.

SVIPT enables to capture single-trial motor performance. Given a high order motor control, the force profile $F(t)$ of a single trial is characterized by a quick force ramp up after the *go-cue* and the avoidance of overshoots. Such an overshoot event is shown in Fig. 3.2. The requested speed-accuracy trade-off can be translated into various performance metrics of the SVIPT task, as described in our conference article [30]. For this presented offline analysis, we selected five metrics to describe single-trial motor performance in SVIPT:

- **Reaction Time / RT**: A quick response upon the *go-cue* is a good start for a successful trial. The time interval between the *go-cue* at time t_{go} and the time point $t_{T0,exit}$, which indicates the cursor leaving the starting field $T0$, is defined as reaction time.

- **Duration / DUR**: Comparable to RT, a short time duration from the *go-cue* at time t_{go} until the hit of the first target field at time point $t_{hit\,1}$ characterizes a successful trial.

- **Cursor Path Length / CPL**: The total path length the cursor is moved from the *go-cue* to the hit of the first target field—named *hit 1*—is described by the integral over the first temporal derivative of the force profile $F(t)$:

$$CPL \equiv \int_{t_{go}}^{t_{hit\,1}} |\dot{F}(t)|\, dt'$$

- **Integrated Squared Jerk / ISJ**: The level of fine-granular motor control is reflected in variations of the trajectory smoothness. Therefore, jerk - defined as the third derivative of the force profile - is expressed by the ISJ metric, which is defined as:

$$ISJ \equiv \int_{t_{go}}^{t_{hit\,1}} |\frac{d^3F(t)}{dt^3}|^2\, dt'$$

- **Normalized Jerk / NJ**: A unit-free variant of ISJ captures smoothness variations. It is given by the *normalized* jerk:

$$NJ \equiv \sqrt{\frac{ISJ \cdot DUR^5}{2 \cdot CPL^2}}$$

Since there are two conditions of target field sequences, a standardization of the performance scores (except for RT) is the prerequisite for pooling trials of both conditions. Therefore, the extracted metrics of each condition were standardized (zero mean and standard deviation one) prior to pooling. Except RT, the metrics are defined with respect to some end point (e.g., $t_{hit\,1}$). Choosing this boundary represents a trade-off between (a) harvesting a metric which is temporally close and thus related to the get-ready interval (the interval before the *go-cue*), and (b) including thorough information about the force trajectory of the current trial. To balance the two conflicting goals, we chose the first hit event *hit 1*.

3.2.4 *Data acquisition and preprocessing*

During a single session, subjects were placed in a chair at 80 cm distance from a 24-inch flat screen. EEG signals from 63 passive Ag/AgCl electrodes (EasyCap) were recorded, which were placed according to the extended 10-20 system. Impedances were kept below 20 $k\Omega$. All channels were referenced against the nose. The signals were registered by multichannel EEG amplifiers (BrainAmp DC, Brain Products) at a sampling rate of 1 kHz. An analog lowpass filter of 250 Hz was applied before digitization. The signal of the force sensor was recorded by an additional amplifier system (BrainAmp ExG, Brain Products).

For outlier identification, the offline preprocessing consisted of low-pass filtering the raw EEG signals at 100 Hz, sub-sampling to 500 Hz sample frequency and high-pass filtering at 0.2 Hz. Therefore, linear Butterworth filters of 5^{th} order were applied. For each trial and all 63 channels, an epoch of 2000 ms duration prior to the *go-cue* was extracted. In order to identify outlier epochs, three rejection methods were applied. First, EEG epochs violating a min-max threshold of 60 μV on frontal channels were excluded from further analysis. Second, a variance threshold on single epochs and channels was applied to remove high-frequent muscular artifacts. Therefore, all cases outside the $[10, 90]$ percentiles and simultaneously exceeding twice the corresponding inter-percentile range were registered as outliers. Third, epochs belonging to extreme trials, represented by outliers of the motor performance metric, were removed. Based on pooled statistics of a single metric, all trials outside the $[20, 80]$ percentiles and also exceeding twice the corresponding inter-percentile range were registered as outliers and removed for further data analysis.The total number of trials N_e entering the following offline analysis procedures varied across subjects and performance metrics. For dataset (D1a), only 2 out of 20 subjects remained with less than 150 out of the original 400 epochs after the preprocessing. We discarded data of these subjects from the following analysis. For dataset (D1b) all 7 subjects remained with at least 150 epochs. The frequency filtering for our main analysis will subsequently be described in Sec. 3.2.6.

3.2.5 *EEG-based single-trial performance prediction*

In this chapter, we aim to predict single-trial SVIPT performance based on pre-trial oscillatory EEG activity within the *get-ready* phase of a single trial. Therefore, the main goal is to gain a neural source estimation $\hat{s}(e)$, whose power achieves the highest correlation with the five SVIPT metrics as continuous variables $z(e)$.

This multivariate regression problem can be tackled with the previously introduced supervised spatial filtering algorithm named SPoC (see Sec. 2.3.4.5). Given N_e bandpassed and epoched multivariate data epochs $\mathbf{X}(e)$ as well as corresponding labels $z(e)$, SPoC learns an optimal spatial filter $\mathbf{w}^* \in \mathbb{R}^{N_c}$ with N_c recorded channels which maximizes the epoch-wise co-modulation between the bandpower of a source $\hat{s}(e)$ and the given target variable $z(e)$.

Applying a SPoC filter $\mathbf{w}_{tr}$ learned from training data $\mathbf{X}_{tr}(e)$, the method allows estimating the target variable z_{est} on novel, unseen test data $\mathbf{X}_{te}(e)$ on a single-trial basis by directly calculating the variance—which approximates the bandpower of the signal—of the narrowband subspace signal:

$$z_{est}(e) = \Phi_{\hat{s}}(e) = \mathrm{Var}[\mathbf{w}_{tr}^\top \mathbf{X}_{te}](e) \overset{\text{Eq. (2.7)}}{=} \mathbf{w}_{tr}^\top \, \mathbf{\Sigma}_{te}(e) \, \mathbf{w}_{tr} \tag{3.1}$$

The last step shows that this subspace variance can simply be obtained by estimating the covariance matrix $\mathbf{\Sigma}_{te}$ on the unseen data epoch $\mathbf{X}_{te}$.

Note, that any subspace components resulting from the SPoC analysis depend mainly on four hyperparameters. In the temporal domain, two of

them define the epoching interval $[t_0, t_0 + \Delta t]$ where t_0 is the starting time relative to the *go-cue* and a duration Δt. In the frequency domain, the lower frequency f_{low} and the bandwidth Δf are the hyperparameters describing the band $[f_{low}, f_{low} + \Delta f]$ to which the data $\mathbf{X}$ is bandpass-filtered.

Even though simple regression of bandpower features on the channel level does not fulfill the requirements of the assumed forward model, we added this simple method for comparison with SPoC. Therefore, channel-wise bandpower features of the training and test set were calculated.

3.2.6 *Evaluation scheme*

Performing a grid search across subjects and SPoC parameters, we restricted the evaluation to a fixed predictive time interval given by $t_0 = -800\,\text{ms}$ prior to the *go-cue* and a window size of $\Delta t = 750$ ms.

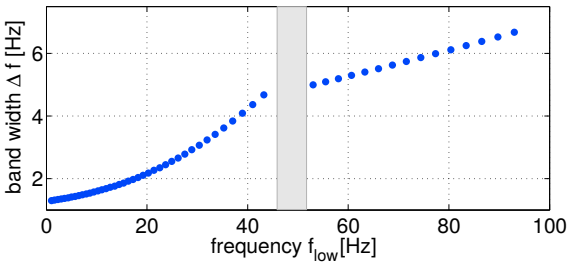


Figure 3.3: **Frequency parameter configurations.** They are characterized by the frequency f_{low} and the corresponding bandwidth Δf. In total, 55 configurations were used for computing SPoC filters. The omitted points (gray area) correspond to the power line frequency range.

As sketched in Fig. 3.3, exponentially increasing and overlapping frequency bands ranging from $\approx$ 1–100 Hz (55 configurations in total) were evaluated from the original non-filtered signals. For bandpass filtering, zero-phase linear Butterworth filters of 5th order were applied. As a trial-wise target variable z, the five different performance metrics introduced in Sec. 3.2.3 were considered. Evaluating SPoC across the complete study group of 18 subjects, using five different motor performance metrics, sweeping through 55 discrete frequency bands and selecting the highest-ranked components (see details below) per configuration, results in more than 12 000 oscillatory components. However, not every single component of a subspace decomposition is expected to be robust, informative and even neurophysiologically plausible [57]. Thus, hereafter we will describe an offline selection strategy in order to identify a subset of the most robust and informative oscillatory components which qualify to predict single-trial motor performance.

For each parameter configuration, a $K = 5$-fold chronological cross-validation (CV) procedure was employed upon the calculation of SPoC [61]. Trials were only considered if they survived the data preprocessing (see Sec. 3.2.4). The N_e EEG pre-trial data epochs $\mathbf{X}$ and their corresponding target variable values z were extracted in chronological order and split

into 5 equally-sized folds. Thus, 4 folds served as training data while the remaining one was used for validating the SPoC filter as described in Eq. (3.1). Since each fold served as test fold once, the estimated target variable $z_{est,j}$ of fold j can be concatenated for all N_e epochs, resulting in $z_{est} = [z_{est,j}]_{j\in[1,5]}$. According to Eq. (2.3), on each fold j the corresponding test pattern is given as $\mathbf{a}_j = \mathbf{\Sigma}_{te,j}\mathbf{w}_{tr}$ utilizing the averaged covariance matrix $\mathbf{\Sigma}_{te,j}$ of the test data $\mathbf{X}_{te}$ in fold j.

The same CV scheme was applied to the linear regression model. The whole parameter space of 3600 configurations was screened. Note, that this number is smaller than the number of components delivered by SPoC analysis, since the latter may deliver more than one component per parameter configuration. The regression, which delivers only a single component per configuration, was trained on the training data and finally applied on test data such that an estimate z_{est} was gained on all N_e trials which survived the data preprocessing step.

For a given parameter configuration, SPoC$_\lambda$ returns a set of N_c filters. As described in [30], it is sufficient to take only the highest-ranked components into consideration[1]. For this purpose, we applied a rank-based criterion: we first removed the linear trend from the ordered set of N_c eigenvalues and obtained a set of N_c residuals r. Only components, which exceeded a threshold of $1.5 \cdot \sigma(r)$ relative to the residual's standard deviation σ, were taken into account for further analysis. We restricted the investigation to positive eigenvalues.

3.2.7 *Evaluation scores to assess the predictive power and model stability*

Given a single component $\mathbf{w}$, we aim to characterize its predictive strength as well as the robustness, e.g., under increased label noise conditions. Therefore, the following set of scores will be considered:

1. **Correlation characteristics:** As a measure to verify the quality of the predictive strength of a SPoC configuration, the overall correlation of the N_e measured performance labels z_{true} with the corresponding predictions z_{est} can be considered:

$$R_{all} = \mathrm{Corr}[z_{true}, z_{est}] \tag{3.2}$$

   Similarly, the predictive strength in terms of single-trial performance can also be verified by checking the mean of the fold-wise correlations $R_j = \mathrm{Corr}[z_{true,j}, z_{est,j}]$, which rewards temporally stable components:

$$R_{folds} = \frac{1}{K}\sum_{j=1}^{K} \mathrm{Corr}[z_{true,j}, z_{est,j}] \tag{3.3}$$

   The correlation based metrics R_{all} and R_{folds} come closest to the original optimization objective of the SPoC algorithm. If the trained

1 The spatial filter solutions are sorted according to their eigenvalues. In case of SPoC$_\lambda$, they equal to the covariance between the bandpower features and the target variable.

spatial filters model trial-to-trial fluctuations well, R_{all} and R_{folds} will report a large value, but only R_{folds} allows discriminating between single-trial predictors and session-trend models. Furthermore, a stable component requires that the correlation of each fold j shares the same sign as the overall value R_{all}. Thus, it is reasonable to require a high homogeneity H_{folds}:

$$H_{folds} = \sum_{j=1}^{K} \Theta(\text{sign}(R_{all}) \cdot \text{sign}(R_j)) \tag{3.4}$$

with $\Theta(x) = 1$ for $x \geq 0$ and $\Theta(x) = 0$ for $x < 0$ representing the unit-step function.

2. **Separability of estimated performance:** Another possibility to characterize the decoding accuracy is to transfer the continuous labels z_{true} into a two-class decoding scenario according to the median of z_{true}. This enables the utilization of the receiver operating characteristic (ROC) curve which is calculated upon the estimated target variable z_{est} given the true two-class labels [141]. As ROC performance can be reduced to a scalar value by calculating the area under the ROC curve (AUC), this novel metric captures the separability of the estimated target variable z_{est} (z-AUC). A perfect decoding corresponds to z-AUC $= 1$ while chance level correspondents to a value of 0.5.

3. **Stressing the stability:** SPoC is a supervised method, which uses label information to guide the spatial filter calculation. Thus, the robustness of a resulting component can be stressed in a quantitative manner by introducing additional label noise. The concept of a stepwise reduction of the SNR of the labels has been introduced in [31]. Here, SNR levels were varied from -20 dB to 10 dB by adding white noise. Applying SPoC, we estimated the target variable z_{est} for all N_e epochs using 5-fold CV. At each SNR level, three sets of noisy labels z were calculated. For each SNR level, the separability of the resulting z_{est} distribution is verified by the z-AUC value. Regarding the z-AUC values as a function of the SNR, the area under this curve — referred to as AAUC_{SNR} — describes the stability of the component (as visualized in the results section by Fig.3.7).

3.2.8 *Selection criteria for robust and predictive components*

To finally identify robust and predictive components, we propose to apply thresholds on three out of the previously introduced five evaluation scores in parallel. As a prerequisite, the dataset was required to consist of at least $N_e = 150$ epochs in order to ensure the convergence of the SPoC algorithm (see [31, 94]):

1. The separability of the predicted performance z_{est} can be verified by the resulting z-AUC value. A corresponding threshold $\text{z-AUC}_{min} =$

0.59 was determined according to the 85^{th} percentile across all configurations.

2. The stability of the component is assessed by the AAUC_{SNR}. Here, a threshold $\mathrm{AAUC}_{SNR,min} = 0.18$ was determined from the 85^{th} percentile.

3. As an additional stability criterion, we require all fold-wise correlations R_j to share equal sign as R_{all} such that $\mathrm{H}_{folds,min} = 5$.

3.3 RESULTS

The results in [28] were gained on the dataset (D1a). Unless noted differently, the results will be shown for dataset (D1a). In the corresponding text, we will comment on components obtained for dataset (D1b).

3.3.1 *SVIPT motor performance scores*

Single-trial based SVIPT performance can be assessed by different metrics (see Sec. 3.2.3). In Fig. 3.4 examples of the trial-to-trial fluctuations of different metrics are visualized for two subjects. The examples cover full sessions, but omit trials removed during the preprocessing. The plots (A) and (D) show the metric reaction time (RT) for the two subjects. It is not affected by a session trend. Its distribution is slightly asymmetric, which is caused by a physiological limit for the minimal RT. The normalized jerk (NJ) in plots (C) and (F) behaves in a similar manner. It is affected only slightly by a global trend, but shows a more skewed distribution compared to RT. In contrast, integrated squared jerk (ISJ) depicted in (B), and cursor path length (CPL) in (E) both show a strong session trend, which can be explained by the user learning (data not shown here). A comparably strong session trend is present also in the duration metric DUR (data not shown).

The cross-correlations between all five metrics and the shape of their distributions were reported in [30]. Metrics ISJ, CPL and DUR showed strong correlations to each other, while RT as well as NJ are both rather independent from the remaining metrics.

3.3.2 *Contrasting SPoC with linear regression on sensor level*

As a baseline comparison for the predictive power of SPoC components, a linear regression model employing channel-wise bandpower features was evaluated as described in Sec. 3.2.8. The resulting distributions of the overall correlation R_{all} and the performance separability z-AUC are reported in Fig. 3.5. Across *all* configurations, SPoC delivers a median correlation $R_{all,med} = 0.07$ and a separability of $\text{z-AUC}_{med} = 0.54$, while on average the regression performs on chance level. While both methods come up with components revealing z-AUC values above chance level, those with the strongest predictive information are generated by the SPoC method.

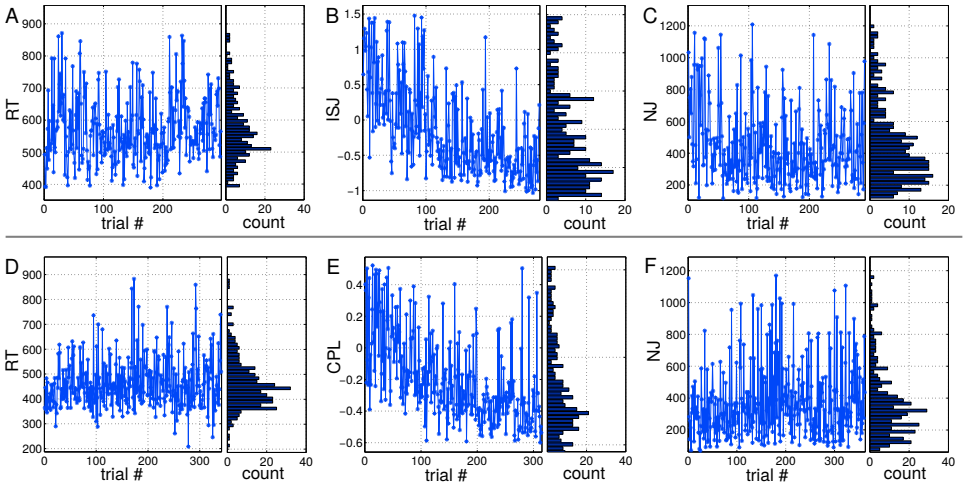


Figure 3.4: **Trial-wise variations of different motor performance metrics.** The examples are extracted over the course of a full session, their histograms are shown in addition. Plots (A)–(C) are taken from data of subject S9, while plots (D)–(F) are from S13.

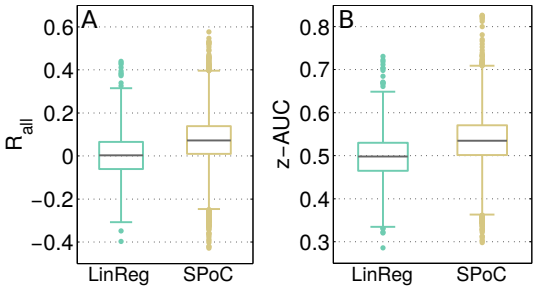


Figure 3.5: **Contrasting the predictive outcome of linear regression on sensor level with SPoC.** Therefore, all 3600 tested parameter configurations for linear regression (LinReg) and over 12000 configurations for SPoC were utilized. In (A), the overall correlation R_{all} between predicted and estimated target variable values is depicted. (B) shows the performance separability z-AUC. Gray lines indicate the median, boxes enclose the 25^{th} to 75^{th} percentile. The whisker length is set to two inter-quartile ranges.

3.3.3 *Single-trial motor performance predictors*

In Fig.3.6, five exemplary predictive and robust SPoC components, gained from five different subjects of dataset (D1a) are characterized. Further examples of dataset (D1b) can be found in A.1. Although SPoC components are computed from band-pass filtered data, the resulting filter $\mathbf{w}$ (gained on all available N_e trials) of a component can be re-applied to non-frequency-filtered epoched data. This spectral content of a SPoC component is shown in (A). The frequency band in which the component was extracted from is indicated by the dashed gray area. Using all available epochs, plot (B) shows the spatial activity pattern gained with Eq. (2.3). In (C), the SPoC filter weights on the 2D-scalp projection are shown. The scatter plot in (D) reports on the measured performance metric z_{true} as a function of the predicted

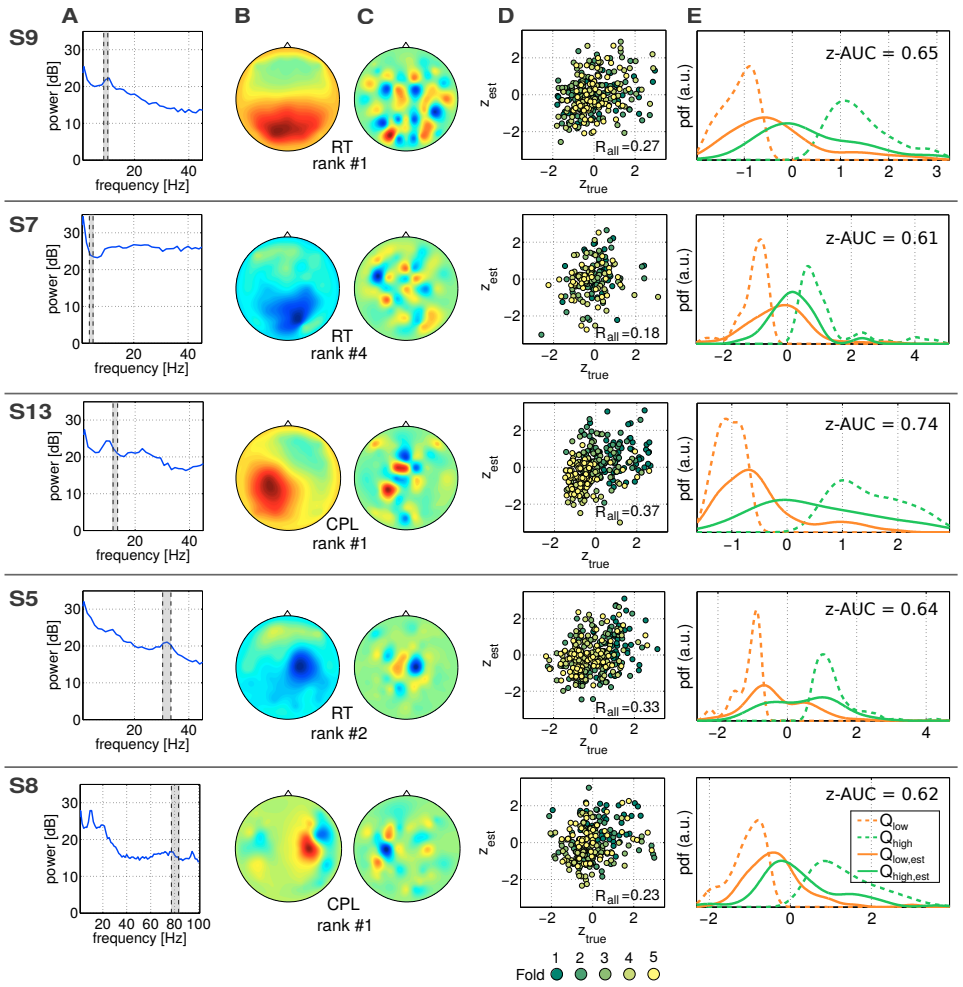


Figure 3.6: **Characterization of *exemplary* predictive SPoC features.** In each row, components are labeled by the used performance metric and the rank according to the full-session filters. (A) Power spectrum of the component applied on non-bandpass filtered full data. The frequency band where the component has been trained is marked by the dashed lines. Note that for the component of S8 a broader frequency range is visualized compared to the other examples. (B) Spatial activity pattern. (C) Filter weights visualized. (D) Scatter plot between true labels z_{true} and the predicted ones z_{est}, color coded by the fold of the chronological CV. (E) To illustrate the separability of the prediction, the distribution of z_{true} values has been split using the corresponding trials of the upper and lower quartiles of z_{est}, which resulted in $Q_{low,est}$ and $Q_{high,est}$. As a reference, the extreme quartiles Q_{low} and Q_{high} of z_{true} are also given (dashed curves). In addition, the z-AUC value based on the median split is reported.

performance z_{est} according to the CV scheme described in Eq. (3.1). The data points are colored by the fold index (1–5), which corresponds to the temporal order of the session. Fold 1 represents the beginning of the session, fold 5 its end. In addition, the overall correlation R_{all} reports on the predictive strength of the component. The distributions shown in (E) illustrate the separability of the single-trial performance values z_{est}. For

this purpose, the estimated labels z_{est} have been reduced to the lower and upper quartile. The corresponding true labels z_{true} were used to compute the quartiles $Q_{low,est}$ and $Q_{high,est}$ and were fitted by a kernel distribution (solid lines). In an ideal case, those quartiles would converge towards the extreme quartiles (Q_{low} and Q_{high}) of z_{true}, which are indicated by dashed lines. As a score of their separability, the z-AUC is given as introduced in Sec.3.2.7.

The exemplary components in Fig. 3.6 are selected across the investigated frequency range depicted in Fig. 3.3. The predictor of S7 can be assigned to the theta band, those of S9 and S13 correspond to the alpha range, the component for S5 originates from the beta range and the one of S8 was found in the gamma range. Regarding the scatter plots, there are two different types of patterns recognizable: single-trial predictors showing a confined point cloud without a clear trend over time (all examples except for S13), whereas the scatter plot of subject S13 shows a clear trend over the course of the session. The separability plots indicate that the predictive power of a single component nicely matches with the z-AUC value.

As an additional analysis, the temporal bandpower dynamics of exemplary predictors on single-trial level can be found in the appendix A.1.

3.3.4 *Testing the stability of SPoC components*

The stability of an oscillatory component can be challenged by systematically reducing the SNR ratio of the target variable z. In Fig. 3.7, the z-AUC score is investigated as a function of the SNR for two parameter configurations. Plot (A) shows a stable component, such that z-AUC is expected to show a step-wise decrease, while for a non-informative component in (B) the z-AUC can be expected to fluctuate around the noise floor. Thus, the resulting *area under the* z-AUC *curve* can be assessed as a tool for mapping the stability of the subspace component under challenging noise conditions. In plot (C), the distribution of this so-called AAUC_{SNR} is reported for all evaluated SPoC components across all 18 subjects. The distribution of AAUC_{SNR} values has its median at 0.07 and is slightly skewed.

3.3.5 *Identification of robust and predictive components*

As described in Sec. 3.2.6, the highest ranked SPoC components of each parameter configuration have been evaluated, resulting in about 12 000 different subspace components for dataset (D1a). In Fig. 3.8, the configurations are characterized by their stability under noise (AAUC_{SNR}), which is plotted in (A) as a function of the separability measure z-AUC, in (B) as a function of the homogeneity of the fold-wise sign of the correlation H_{folds} and in (C) as a function of the overall correlation R_{all}. A few observations can be made: First, the metrics are not centered at zero. Second, based on all initial configurations (blue data points), AAUC_{SNR} correlates with the z-AUC as well as with R_{all}. The largest AAUC_{SNR} values are evoked by the most homogeneous fold-wise correlation signs with $H_{folds} \geq 3$. The thresh-

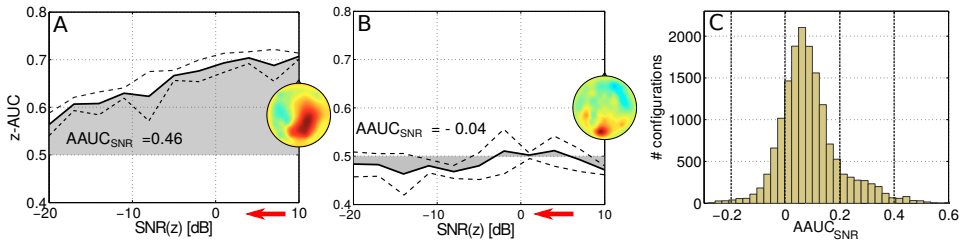


Figure 3.7: **Stressing the stability of two exemplary SPoC components.** For two different parameter configurations (A) and (B), the z-AUC-values (solid lines) describing the separability of the prediction are plotted together with standard deviations (dashed lines) for a stepwise decrease of the SNR ratio (indicated by the red arrow). The area under the z-AUC curve — further on called AAUC_{SNR} — describes the stability of the component under the challenge of added noise. Plot (C) shows the histogram of all AAUC_{SNR} scores evaluated for the considered parameter configurations.

old criteria applied to select the best of the 12 000 subspace components are indicated by red dashed lines, and red dots indicate the finally selected components. Here, only components with a consistent sign of R_{folds} on the test data were selected.

As shown in Fig. 3.9 (A) and (B), the overall correlation R_{all} is strongly correlated with the z-AUC metric, such that an additional threshold criterion on R_{all} was not necessary. For dataset (D1a), the most predictive components achieve a correlation value of up to 0.6, corresponding to ${R_{all}}^2 = 0.36$. Assuming a linear relationship between z_{true} and z_{est} as well as normally distributed data, this means that z_{est} can explain up to 36 % of the performance variance contained in z_{true}. For dataset (D1b) with generally less training data, best predictors achieve correlations up to $R_{all} = 0.4$.

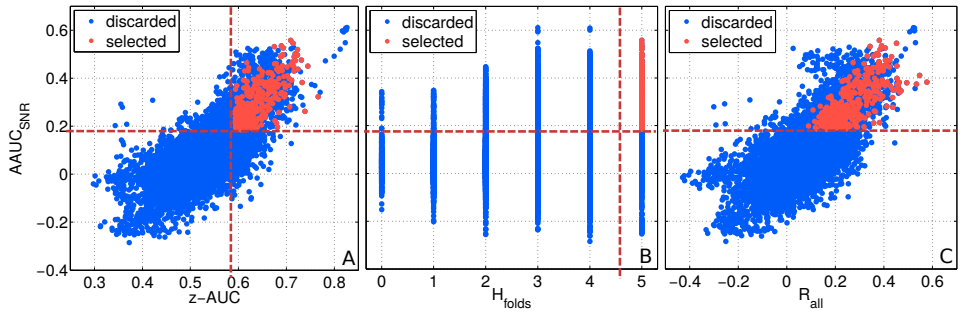


Figure 3.8: **Characterizing the space of SPoC components in terms of their stability and predictive information.** The SNR-challenged AAUC_{SNR} is given as a function of the performance separability z-AUC (A), in relation to the homogeneity of the correlation sign H_{folds} (B), and dependent on the overall correlation R_{all} (C). Red data points describe the selected SPoC components after applying thresholds (dashed red lines).

In Fig. 3.10, all 361 selected components for dataset (D1a) and 140 for (D1b) are characterized by histograms in terms of their input parameters.

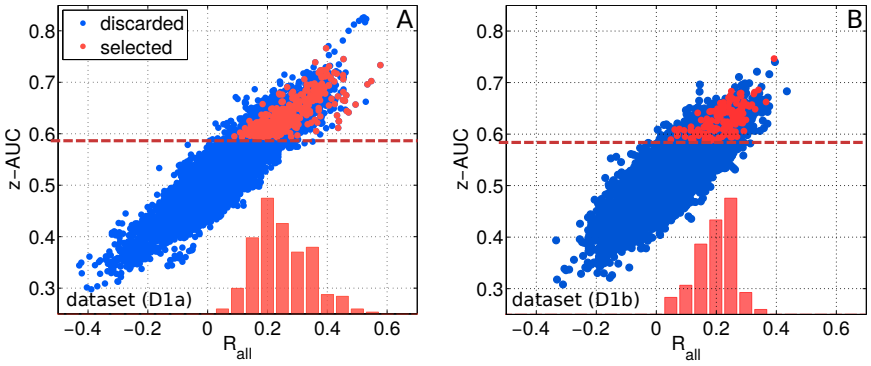


Figure 3.9: **Relation between separability score and the overall correlation for both datasets.** (A) shows results of dataset (D1a), while (B) reports it for (D1b). In both plots, all computed components (blue dots) and the selected ones (red dots) are shown. The dashed red line indicates the threshold z-AUC$_{min}$ applied to select the most informative components. The red bars indicate the distribution of R_{all} values for the selected components only.

(A) and (E) display the subject-wise grouping. In total, 16 out of 25 subjects contribute at least one component, for three subjects more than 50 configurations survive the selection procedure. (B) and (F) characterize the selected components assigned to their underlying frequency band $[f_{low}, f_{low} + \Delta f]$ (see Fig. 3.3). For both datasets, most components are gained from the alpha- and beta-band range. Interestingly, robust features detected in the gamma band were dominantly selected for their ability to predict CPL. The slow frequency (<4 Hz) components are dominated by artifactual subspaces. (C) and (G) report on the occurrences of the different performance metrics among the selected components. Most components could be extracted for RT (61 %), followed by CPL (14 %). (D) and (H) provide an overview over the SPoC ranks of the selected components. The rank ordering corresponds to the eigenvalue ordering of the complete dataset. As the number of selected components drop with increased rank, the ranking is associated with the information content of the subspace component.

SPoC provides linear spatial filters that allow for a limited but still important neurophysiological interpretation of spatial activity patterns. A representative subset of typical scalp topographies from the selected stable and informative subspaces are plotted in Fig. 3.11. The components were assigned to three groups. About 70 % of components fall into group G1, which comprises patterns ranging from activations in occipital, to central or frontal areas. The maximum activity of those components often is found over one of the hemispheres. About 10 % of the components fall in group G2. They show patterns of probable non-neural sources and may represent, e.g., eye artifacts, muscular activity or single noisy channels. Group G3 comprises noisy topographies. As indicated by patterns in the intersection area of the three groups, mixed components were observed as well.

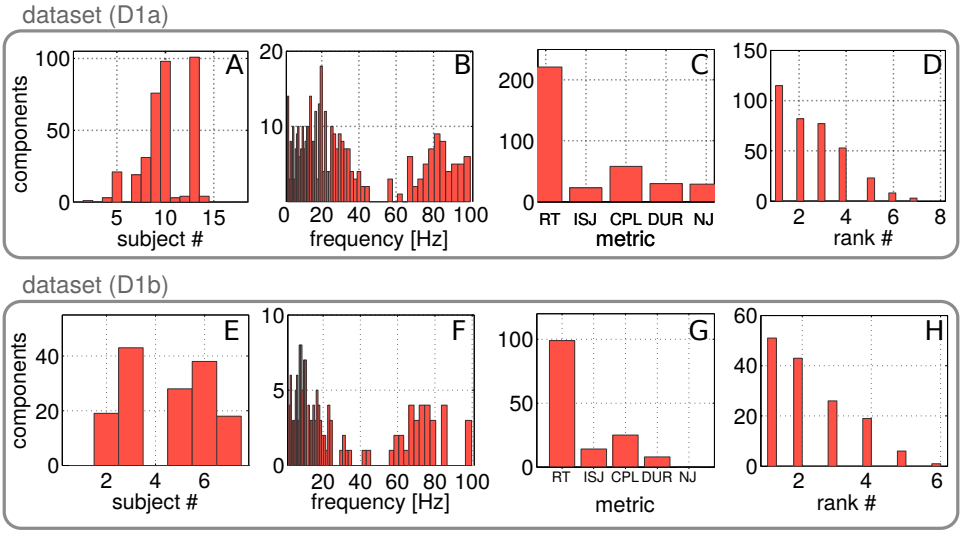


Figure 3.10: **Histograms of involved hyperparameters solely for the selected SPoC components.** The first row reports the distributions for dataset (D1a), the second row for (D1b). (A) and (E) reveal the assignment to the single subjects. (B) and (F) visualize the distribution across frequency bands. (C) and (G) depict the spread of components over the five utilized motor performance metrics. (E) and (H) show the split according to their SPoC rank positions.

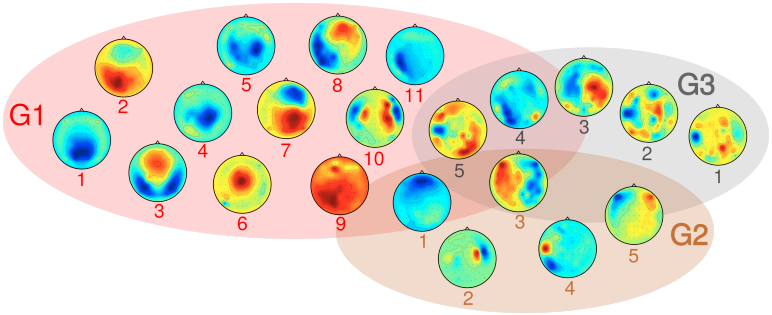


Figure 3.11: **Overview of typical activity patterns.** The selected components were grouped in three categories: G1 consists of components with neural origin, G2 comprises artifact-related subspaces and G3 captures non-informative components.

3.4 DISCUSSION

We hypothesized that subject-specific pre-trial brain signals contain information which allows us to partially explain and temporally predict the trial-by-trial variability of upcoming motor performance in SVIPT. To test the hypothesis, we developed a workflow which is capable to extract informative oscillatory EEG subspace components and to identify the most robust ones. In an online simulation, our analysis revealed strong evidence that the bandpower of the selected components is predictive for the upcoming single-trial SVIPT performance. Major findings were that these components indeed exist, but need to be optimized for individual users. With 16 out of 25, not all, but a majority of the subjects revealed the desired

robust and informative features. As reported in [142], our findings on data of seven chronic stroke patients suggest that the workflow also allows extracting robust motor performance predictors under more severe conditions, such as less training trials and intensified artifactual contributions. In the following, we will first discuss the decision to utilize SPoC instead of other alternative analysis methods. In this context, the proposed selection procedure and the stability of SPoC components over time is discussed, with a special focus onto the role of SNR, frequency and the BCI illiteracy phenomenon. In addition, the detected components will be related to existing literature and characterized on a group-level with respect to the covered frequency bands, sub-processes reflected by the components and the time courses revealed. Before concluding, we will describe a neuroergonomically enhanced rehabilitation paradigm as a possible use case of our contribution.

3.4.1 *SPoC and its alternatives*

Designing the data analysis workflow, we built upon our background in BCI. Accordingly, we carefully selected algorithmic building blocks only, if they can be applied in single-trial analysis (e.g., the application of the spatial SPoC filter according to Eq. (3.1)). This decision should simplify the translation of the presented workflow to closed-loop experiments, as carried out in the final Chap. 6. The choice of the supervised SPoC algorithm for extracting informative components is supported by its good performance compared to a supervised linear regression of bandpower features on the sensor level (see Sec. 3.3.2). This is in accordance with findings of Dähne et al. [94]. On data from an auditory steady-state evoked potential paradigm, the authors reported better results for SPoC compared to both, linear regression and an ICA decomposition. SPoC does not reconstruct sources of the brain, but instead performs a supervised subspace decomposition and delivers discriminative information. Thus, a SPoC subspace component cannot be expected to correspond to a single physical source or even a dipole source (even though such SPoC components are possible). Theoretically even several spread-out brain areas may contribute to a single SPoC component, if they share oscillatory activity which co-varies over time with the labels. The choice between SPoC and source reconstruction approaches [124] represents a trade-off — while the latter may facilitate the interpretation of results, SPoC components avoid several of the drawbacks mentioned in Sec. 2.3.2. As our workflow was aligned in terms of applicability for single-trial online paradigms, our decision was biased towards SPoC.

In principle, a preceding dimensionality reduction step, e.g., by using a PCA decomposition, before the training of the SPoC model could have been performed. However, as recently reported by [143], such a preprocessing step might negatively affect the stability and quality of resulting oscillatory subspace components. Thus, we have spared this ingredient for the introduced workflow.

3.4.2 *Selection criteria for robust and predictive components*

Over-fitting is a general issue for supervised methods and for SPoC in particular, as no form of regularization was applied. This requires some form of post-hoc selection of SPoC components. The situation is aggravated, as SPoC returns full rank filter matrices, which result in a very large numbers of subspaces. However, only a fraction of these can be expected to be informative about the labels. As robustness over time as well as with respect to label noise are important criteria for the potential closed-loop applicability of a component, a single selection criterion (e.g., a threshold on the correlation value) is not sufficient. By that, we selected three criteria (see Sec. 3.2.8), which suited best these requirements. Out of the initial five selection criteria, the two scores R_{all} and R_{folds} turned out to be beneficial for characterizing the extracted components. Thus, they were omitted for the selection process, since a strong correlation between z-AUC and R_{all} was observed (see Fig. 3.9). The same holds for the correlation between z-AUC and R_{folds} (not shown). An alternative to this selection procedure would be to relax the thresholds and combine it with additional methods to judge the plausibility of the remaining components post-hoc. For ICA components, workflows have been proposed, such as MARA, an automatic classification of artifactual components by [89]. MARA uses features based on topology, time-frequency analysis and source reconstruction. Similar approaches have been proposed by [144] and [145].

3.4.3 *Influence of SNR on SPoC components*

By applying rather strict selection criteria, weaker but still informative components may have been removed. As a result, the data of some subjects did not reveal informative pre-go oscillatory components. This characteristic might be due to a lower SNR of their data, which hides potential informative content from the SPoC analysis, especially in combination with the limited number of trials used. The work of [31] on robustness testing of SPoC components backs this interpretation. In this case, future improvements may be expected by regularization techniques introduced to SPoC (see Chap. 4) or from transfer learning approaches [146]. However, we cannot exclude that informative oscillatory components may not be visible to the EEG or may be absent in some subjects. This problem has been described as BCI "illiteracy". It has predominantly been studied in the context of motor imagery paradigms for the control of BCI applications [147], in which the decoding of the imagery class usually is not possible for a subset of subjects. The BCI illiteracy problem was tackled by novel experimental setups like hybrid BCI paradigms [148, 149], but could also be alleviated by more advanced decoding methods [150].

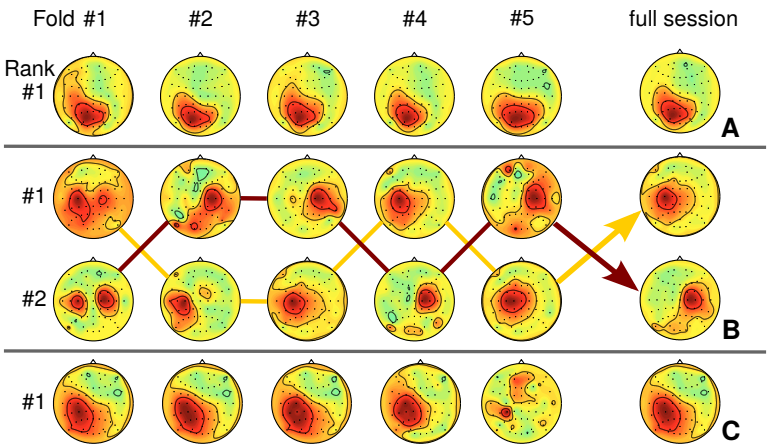


Figure 3.12: **Relation between SPoC rank stability and pattern homogeneity.** The analysis was performed over five cross-validation folds (chronological order). (A) Stationary case: component is first-ranked across all five folds (data of subject S9, $f = [9.4, 11]$ Hz, RT). (B) Rank switching: Two almost stable components switch rank positions between folds (S5, $f = [27.5, 30.3]$ Hz, RT). Lines connect the corresponding topologies. (C) Intensity variation: intensity of first-ranked component decreases over time folds (S13, $f = [13.6, 15.3]$ Hz, CPL)

3.4.4 *Rank stability of SPoC components over time*

In Sec. 3.4.3, the relation between SPoC solutions and the SNR of the data has been touched. As SPoC ranks the detected components according to their covariance values with the true metric, solutions may seem unstable when only the first-ranked component is considered. In real-world datasets, variations of the SNR over time can induce rank switches or mixed components. Tracking a component over multiple runs of the subspace decomposition method is a challenging task, especially as mixtures theoretically can not be distinguished from a single source. However, as similar problems arise for online learning of blind source separation methods like ICA, practical solutions are available [151]. Fig. 3.12 gives examples of stable, stationary components (A) and of unstable SPoC components (B) and (C), both observed over the five chronological CV folds. Instable components may be evoked if the stationarity assumption of SPoC is violated, e.g., by slow temporal intensity variations due to user learning. For (B), arrows indicate a possible path through rank positions across folds by connecting corresponding components. Please observe, that SPoC generates cases with even more severe variation between folds as those depicted in (B) and (C). However, such components typically have been removed during the selection process. While mixed, yet stable components may be hard to interpret, they can still be useful for predicting the task performance.

We have observed a high sensitivity of SPoC for small differences in the frequency parameters. Seemingly unstable components which display rank switching behavior (see Fig. 3.13 at 8.7 Hz) can sometimes be stabilized by slightly changing the frequency, e.g., to 9.4 Hz in this example.

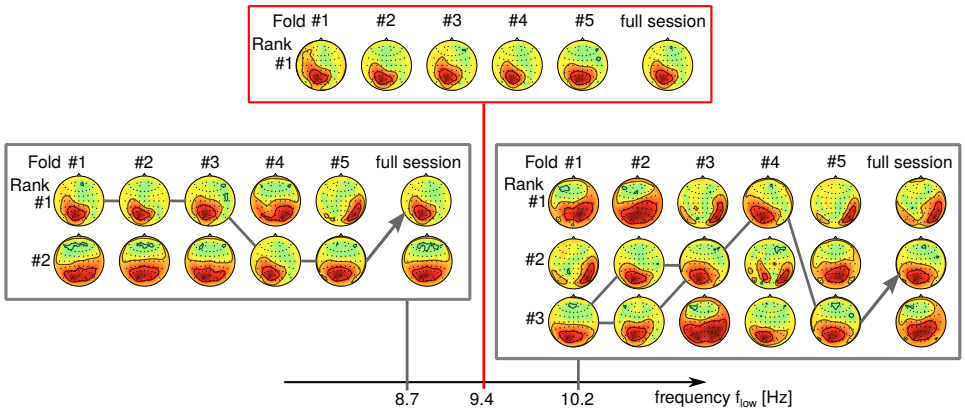


Figure 3.13: **Influence of the frequency band upon the rank stability.** Based on data of subject S9, we find a stable component at $f_{low} = 9.4$ Hz, while it develops rank instabilities with a slight variation of the frequency band.

Further increase of the frequency to 10.2 Hz again induces instability in this example.

Focusing on further introspection of oscillatory components, we have also proposed a post-hoc sensitivity analysis of a single SPoC filters [152, 153]. After estimating a SPoC filter under a specific choice of time-frequency parameters, the filter was applied on data of broadly defined time-frequency hyperparameter space. Such a post-hoc hyperparameter sensitivity analysis can provide additional introspection about the component's persistence, such as identifying the timescale and frequency range in which this specific oscillatory subspaces can explain or even predict label information.

3.4.5 *Characterization of robust SPoC components and sub-processes*

The proposed SPoC workflow delivers a diverse set of oscillatory components, which vary in their topological patterns as well as in their underlying frequency band. This is not surprising, since SVIPT requires the interaction of several cognitive sub-processes in order to reach a good overall performance. For each sub-process, one or more specific neural features may exist, with all of them being informative about the overall outcome of the complex task.

The best components differ between subjects and predominantly occur in the alpha band and beta band. Our findings are supported by informative features in the alpha and beta-range observed during pre-movement intervals of a hand grasping task [134, 138, 139]. Furthermore, the informative frequency ranges for SVIPT are comparable to those reported for attention related tasks [124, 125, 154]. We obtained best results when using RT as a performance metric, which supports our earlier findings on disjunct data from younger subjects [29, 155]. RT of course does not automatically lead to a successful trial, but it can be seen as an indicator for a quick ramp-up phase and alertness. For fewer users, presumably those with highest SNR

characteristics, informative oscillatory features could be identified for other performance metrics of the force task, too.

Comparing the topological plots of group G1 in Fig. 3.11 with those reported in literature, it can be observed that many of them resemble patterns emerging for motor imagery tasks in BCI [156]. These often display a clear maximum of activity in channels located over one of the sensorimotor areas (compare pattern 5 of Fig. 3.11 and the pattern of S5 in Fig. 3.6) or are located centrally over both hemispheres. While similarity of patterns are by no way a proof for an origin of these oscillatory components in the sensorimotor cortices, the hand force action required to succeed in the SVIPT task would allow for such components.

Other components show a maximum intensity over parietal and occipital areas and may reflect the involvement of the visual system in the SVIPT task. Pattern 2 of Fig. 3.11 and all patterns in Fig. 3.12 (A) display a lateralization similar to patterns reported for *directed* and covert visual attention processes [121, 157]. Components with a centrally located maximum (compare pattern 1 of Fig. 3.11 or the pattern of S9 in Fig. 3.6), or with double wing shapes (e.g., pattern 3 in Fig. 3.11) resemble components reported for generalized visual attention processes [46, 139]. Again, most of these rather clear patterns originate from the alpha frequency band.

While the relevance of several of the selected components cannot be fully interpreted, we do consider these features as added value for neurologists, e.g., by tracking the power time course over sessions for a subject-specific component. Further insight into underlying sub-processes and participating brain areas may be obtained from a post-hoc source reconstruction applied upon single SPoC subspaces.

3.4.6 *Behavioral variability on different time scales*

Independent of the choice of the exact motor task, subjects generally display two types of performance variations [158]. First, a large trial-to-trial performance variability is observed from behavioral data. Second, slow performance drifts can occur over the course of a session. Accordingly, SPoC can deliver components, which reflect either one of the two types of performance variations. To tell them apart, a comparison between R_{all} (or the strongly correlated metric AAUC_{SNR}) and R_{folds} is helpful. High values for R_{all}, but low ones for R_{folds} indicate a session trend. If both are high, then the component is informative for trial-by-trial variation (see *single-trial predictors* and *session trend predictors* in Fig. 3.14 as well as the examples given in Fig. 3.6).

For the purpose of brain state-informed closed-loop experimenting, single-trial predictors may be more suitable. Session trend predictors, however, may still be useful for pre-cleaning the performance labels. While session trend predictors may reflect an increasing fatigue or a learning effect, it is much harder to determine underlying mechanisms, which cause the rapidly changing trial-to-trial performance of the single-trial predictors [24, 26, 111]. However, our identified components reveal strong

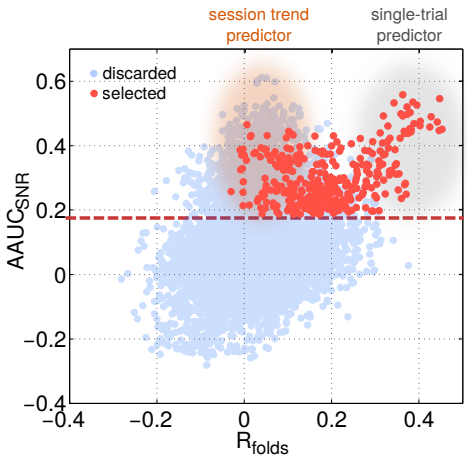


Figure 3.14: **Identification of session trends vs. single-trial performance variations.** This information becomes accessible by inspecting a predictor's characteristic with respect to two selection criteria. The scatter plot visualizes AAUC_{SNR} as a function of the mean correlation value across folds R_{folds} for all configurations (light blue) and the selected ones only (red). Two classes of predictors can be identified: single-trial predictors showing a high R_{folds} value while session-trend predictors show a very low R_{folds} value.

evidence that the pre-trial brain activity is partially informative about trial-by-trial variability of motor performance. This finding is in accordance with [159] who reported on monkey experiments that at least 30 % of behavioral variability could be explained by the fluctuations of preparatory neural activity in the dorsal premotor cortex. However, [158] stated only a weak relationship between motor cortex activity (PMd/M1) in monkeys and trial-wise fluctuations of behavior.

3.4.7 *Outlook: closed-loop experimenting*

The predictive EEG features were extracted from a pre-go interval of each trial. Our pipeline carefully simulated an online scenario, but this approximation of course can not replace the evaluation within an online study. The exploitation of the identified individual informative features in a closed-loop experiment will finally be provided in chapter 6. Beforehand, the two following chapters will first tackle the robustification of the utilized decoding method in order to enhance the feasibility of the final application with stroke patients.

3.5 LESSONS LEARNED

In summary, we introduced a workflow to identify subject-specific single-trial based neural markers which are predictive for the performance of an upcoming motor task. In an online simulation, we were capable to identify such predictors and tested their robustness both on data of normally aged controls as well as on chronic stroke patients. Such robust predictors can

be valuable building blocks for closed-loop applications since they provide introspection about, e.g., sub-processes involved in hand motor control, and can directly be applied in single-trial experimenting at low computational effort. This is of specific interest for brain state-informed paradigms, e.g., in post-stroke rehabilitation. Furthermore, the group-level analysis motivated to utilize our workflow to gain a better global understanding of trial-to-trial variations of cognitive sub-processes, which could finally support a successful rehabilitation outcome.

4

REGULARIZATION TECHNIQUES FOR SPATIAL FILTER OPTIMIZATION

This chapter mainly refers to the journal paper "Characterizing Regularization Techniques for Spatial Filter Optimization in Oscillatory EEG Regression Problems" by Meinel et al. [32]. Regarding this collaborative project, I have taken the lead in developing and implementing a number of potentially beneficial regularization strategies, in developing evaluation strategies to judge their efficacy, in visualizing the results and in writing the manuscript. In addition to the content of this paper, two more real-world datasets have been included to the analysis.

SUMMARY

Novel supervised algorithms for single-trial brain state decoding are reported. Their reliability and robustness are essential to efficiently perform neurotechnological applications in closed-loop. When brain activity is assessed by multichannel recordings, spatial filters computed by the source power comodulation (SPoC) algorithm allow identifying oscillatory subspaces. They regress to a known continuous trial-wise variable reflecting, e.g., stimulus characteristics, cognitive processing or behavior. In small dataset scenarios, this supervised method tends to overfit to its training data as the involved recordings via EEG, MEG or local field potentials generally provide a low signal-to-noise ratio. To improve upon this, we propose and characterize three types of regularization techniques for SPoC: approaches using Tikhonov regularization (which requires model selection via cross-validation), combinations of Tikhonov regularization and covariance matrix normalization as well as strategies exploiting analytical covariance matrix shrinkage. All proposed techniques were evaluated both in a novel simulation framework and on two real-world datasets. Based on the simulation findings, we saw our expectations fulfilled, that SPoC regularization generally reveals the largest benefit for small training sets and under severe label noise conditions. Relevant for practitioners, we derived operating ranges of regularization hyperparameters for cross-validation based approaches.

4.1 INTRODUCTION

4.1.1 *Motivation*

Neurotechnological systems such as BCIs typically utilize EEG recordings that enable users to interact with a computer or physical devices [8]. Such practical closed-loop applications require the extraction of relevant and robust features [160] from high-dimensional EEG data which unfortunately suffer from an inherently low signal-to-noise ratio [18, 161]. In addition, for most BCI applications only small calibration datasets are available to train the decoding algorithms—typically a few dozens or maximally a couple of hundreds of training samples — which further aggravates the situation [62]. Thus, it is necessary to design robust decoding methods and training procedures, such that over-fitting to the training data is avoided [18]. In particular, these arguments also apply for the SPoC algorithm which tackles the single-trial decoding of a continuous variable (see the previous Chap. 3).

While robust variants of CSP have been proposed based on regularization approaches [93], there are no such robust variants for SPoC. Comparing the formulations of the objective functions of SPoC and CSP (see Sec. 2.3.4.4 and 2.3.4.5), both can be translated into a similar optimization problem.

Research question Q2a

Thus, we present generally applicable regularization variants for SPoC and evaluate if the algorithm's robustness can be increased. Therefore, we first evaluate regularized SPoC variants using a recent simulation approach based on post-hoc labeling of arbitrary EEG recordings. This allows probing the stability of the regularized variants under reduced training datasets, varying label noise conditions and different strengths of oscillatory sources. In a second step, we evaluate the methods on two different real-world datasets and finally compare the findings of both scenarios. As regularization introduces additional hyperparameters, we will compare model selection via cross-validation with an analytical solution. Finally, we provide the practitioner with information on how to determine suitable parameters for SPoC regularization.

4.1.2 *Related work*

In the BCI field, the most prominent algorithm for a supervised classification scenario is tackled by the CSP algorithm (see Sec. 2.3.4.4). Unfortunately, CSP is specifically sensitive towards noisy training data [162], non-stationarities [163] and small datasets [164, 165]. To mitigate a subset of these limitations, regularization variants have been proposed for CSP [93, 163]. In general, regularization guides an optimization problem by adding prior information, thus limiting the space of possible solutions. Even though regularization is of specific importance for ill-posed problems such as source reconstruction [166], less underdetermined problems can also profit. For CSP, a broad bandwidth of regularization approaches has been published, such as L1- and L2-norm penalties [93, 167–169], regularized transfer learning strategies that accumulate information across multiple

sessions and subjects [170–174] and variants which favor invariant solutions across sessions/runs under EEG non-stationarities [163, 175–177].

Taking a closer look into the BCI decoding literature, a variety of methods for oscillatory EEG classification problems can be found, but for the regression case the choice still is extremely limited [178] even though regression methods allow tackling highly interesting research questions. Examples are the estimation of continuous mental workload levels [179, 180], decoding the depth of cognitive processing [181], predicting single-trial motor performance [28] or continuous decoding of movement trajectories [182]. Thus, we will focus on regularization variants for the SPoC algorithm in the following chapter.

4.2 REGULARIZATION FOR REGRESSION BASED SPATIAL FILTERING

The constraint optimization problem of the SPoC algorithm was introduced in Sec. 2.3.4.5. It can be translated into the following Rayleigh quotient:

$$J_\lambda(\mathbf{w}) = \frac{\mathbf{w}^\top \mathbf{\Sigma}_z \mathbf{w}}{\mathbf{w}^\top \mathbf{\Sigma}_{avg} \mathbf{w}} \tag{4.1}$$

In most BCI scenarios small training datasets—typically less than 100 samples—of a high dimensionality are encountered [18]. In this setting, SPoC shows an impairing sensitivity and thus might be prone to overfit the training data [31]. A common machine learning strategy in such situations is to add prior information and thus regularize the objective function of an algorithm.

Similar to the regularization strategies proposed by Lotte and Guan [93] for CSP, there are two possible branches of regularization strategies for the SPoC algorithm: The first is to directly add prior information on the level of the objective function in Eq. (4.1). This leads to a restriction of the solution space of possible filters. The second one directly addresses the involved empirical covariance matrices which suffer from small training sets and noisy data. Poorly estimated covariance matrices will not characterize the intended neural activity well. Therefore, regularization on the level of covariance matrices intends to improve their estimation and thus enhance the spatial filtering optimization. In the following, we will propose two regularization approaches, one from each branch of strategies.

4.2.1 *Additional penalty on the objective function*

Introducing a regularization to the objective function of SPoC can be achieved by adding a penalty term $P(\mathbf{w})$ to the denominator of the Rayleigh quotient stated in Eq. (4.1):

$$\widetilde{J}_\lambda(\mathbf{w}) = \frac{\mathbf{w}^\top \mathbf{\Sigma}_z \mathbf{w}}{(1-\alpha)\mathbf{w}^\top \mathbf{\Sigma}_{avg} \mathbf{w} + \alpha P(\mathbf{w})} \tag{4.2}$$

where $\alpha \geq 0$ is the regularization parameter that modulates the strength of the penalty. In general, the term $P(\mathbf{w})$ penalizes solutions of $\mathbf{w}$ that do not

fulfill a specified prior. Thereby it increases the algorithm's robustness to outliers and small training sets.

Here, we select a simple quadratic penalty of the form:

$$P(\mathbf{w}) = \mathbf{w}^\top \mathbf{I}\, \mathbf{w} = \|\mathbf{w}\|^2 \tag{4.3}$$

using the identity matrix $\mathbf{I} \in \mathbb{R}^{N_c \times N_c}$. This penalty is known as *Tikhonov regularization* [183] and has similarly been introduced for CSP [93]. As the penalty $P(\mathbf{w})$ scales with the spatial filter norm, solutions $\mathbf{w}$ with small weights are preferred. Regarding utmost regularization strength in Eq. (4.2) expressed by $\alpha = 1$, the Rayleigh quotient simplifies to the one of PCA (see Sec. 2.3.4.1) meaning that a PCA on the z-weighted covariance matrix is computed. For the introduced Tikhonov regularization of SPoC, model selection wrt. α can be done via cross-validation (CV).

Equivalence to covariance shrinkage

Inserting the given Tikhonov penalty $P(\mathbf{w})$ of Eq. (4.3) into the objective function in Eq. (4.2), enables to factorize the denominator to a shrinkage of the averaged covariance matrix $\mathbf{\Sigma}_{avg}$ towards the identity matrix $\mathbf{I} \in \mathbb{R}^{N_c \times N_c}$:

$$\widetilde{\mathbf{\Sigma}}_{avg} = (1-\alpha)\mathbf{\Sigma}_{avg} + \alpha \mathbf{I} \tag{4.4}$$

By that we have shown that substituting $\mathbf{\Sigma}_{avg}$ by the shrinked version $\widetilde{\mathbf{\Sigma}}_{avg}$ in the objective function of SPoC (see Eq. (4.1)) is equivalent to the Tikhonov formulation stated in Eq. (4.2) and (4.3).

Trace normalization

SPoC$_\lambda$ optimizes covariance, which is not scale-invariant. This drawback might be mitigated by the norm constraint, but to directly compensate for the relative scaling of the covariance matrices in Eq. (4.2), a normalization of *all* covariance elements by the trace $tr[\cdot]$ might also be a suitable strategy as already proposed for CSP [92, 184]:

$$\widehat{\mathbf{\Sigma}}(e) = \frac{\mathbf{\Sigma}(e)}{tr[\mathbf{\Sigma}(e)]}; \quad \widehat{\mathbf{\Sigma}}_{avg} = \frac{\mathbf{\Sigma}_{avg}}{tr[\mathbf{\Sigma}_{avg}]} \tag{4.5}$$

Here, we investigate the effect of applying trace norm to $\mathbf{\Sigma}(e)$ and $\mathbf{\Sigma}_{avg}$ entering Eq. (4.2), but not upon $\mathbf{\Sigma}_z$ as its label-weighting shall be maintained.

4.2.2 *Regularization of covariance matrices*

In parallel to the proposed Tikhonov regularization, which builds upon a CV procedure for the selection of α, there are faster ways of determining a suitable regularization strength. Hereafter, we will focus on two strategies for covariance shrinkage which allow to deploy an analytic solution to determine the regularization parameter.

Automatic shrinkage of sample covariance matrices

In general, when estimating a sample covariance matrix $\mathbf{S} \in \mathbb{R}^{N_c \times N_c}$ based on N_{train} training data samples, there was a systematic bias reported if N_{train} is in the order of the given dimensionality N_c or yet below: large eigenvalues get overestimated while small eigenvalues tend to be underestimated [185]. The situation can be improved by shrinking the covariance matrix S towards the identity matrix $\mathbf{I}$ [186, 187]:

$$\widetilde{\mathbf{S}} = (1 - \alpha^*)\mathbf{S} + \alpha^* \nu \mathbf{I} \tag{4.6}$$

Under the assumption of independent and identically distributed (i.i.d.) data and thus in the absence of outliers, Ledoit & Wolf derived a closed-form solution for the optimal shrinkage parameter α^* and the optimal scaling parameter ν by minimizing the expected mean squared error. For the exact closed-form solution of α^* and ν, we refer the reader to Ledoit and Wolf [186], Schäfer and Strimmer [187], and Bartz and Müller [188]. This closed-form solution holds the advantage of directly computing an estimate of α^* without cross-validation. Note that the additional scaling factor ν takes a similar role as the trace normalization introduced for Tikhonov regularization with the difference that it only takes diagonal terms into account.

Automatic shrinkage of averaged covariance matrix

As shown in Sec. 4.2.1, the Tikhonov penalty introduced for the SPoC objective function can be rewritten as a shrinkage of the covariance matrix $\mathbf{\Sigma}_{avg}$, which was gained by averaging across the epoch-wise covariances $\mathbf{\Sigma}(e)$. Thus, one can directly apply the closed-form solution for α^* and ν, but it first requires to estimate the averaged covariance matrix as $\mathbf{\Sigma}_{avg} = (N_s \cdot N_e - 1)^{-1} \mathbf{X}_{cat}^\top \mathbf{X}_{cat}$ using a concatenated data matrix of all N_e epochs, namely $\mathbf{X}_{cat} = [\mathbf{X}(1), ..., \mathbf{X}(N_e)] \in \mathbb{R}^{N_c \times (N_s \cdot N_e)}$. To compensate for signal non-stationarities, each data epoch $\mathbf{X}(e)$ should be corrected to channel-wise zero mean prior to concatenation.

Automatic shrinkage of epoch-wise covariance matrix

SPoC includes the label-weighted covariance matrix $\mathbf{\Sigma}_z$ which holds all the available label information. A direct covariance shrinkage for $\mathbf{\Sigma}_z$ was tested in pilot experiments (data not shown), but this turned out not to be beneficial — probably because adding a regularization term would diminish the contained label information. As both $\mathbf{\Sigma}_z$ and $\mathbf{\Sigma}_{avg}$ require a computation of the epoch-wise covariance $\mathbf{\Sigma}(e)$, we propose to choose this matrix as regularization target using the previously mentioned closed-form solution for α^* and ν in order to derive a shrinked estimate $\tilde{\mathbf{\Sigma}}(e)$.

4.2.3 *Overview of evaluated SPoC regularization variants*

Tab. 4.1 gives an overview of all proposed regularization strategies which were introduced in Sec. 4.2.1 and 4.2.2. The first three rows summarize Tikhonov regularization variants which all require an estimation of α by means of cross-validation. Among them, **Tik-SPoC** comprises Tikhonov regularization only according to Eq. (4.2), while **NTik-SPoC** considers an additional trace norm both for $\mathbf{\Sigma}(e)$ and $\mathbf{\Sigma}_{avg}$. The largest extent of regularization is realized by **ASNTik-SPoC** which uses the same strategy as NTik-SPoC with additional automatic shrinkage on $\mathbf{\Sigma}(e)$ for the computation of $\mathbf{\Sigma}_z$. As this term applies to the numerator (N) of the objective function, this is marked accordingly in Tab. 4.1. The last two rows summarize automatic shrinkage approaches using the closed-form solution by Ledoit & Wolf (LW). Applying automatic shrinkage to the averaged covariance matrix will be referred to as automatic Tikhonov regularization **aTik-SPoC**. In contrast, using automatic shrinkage directly upon $\mathbf{\Sigma}(e)$ in the numerator (E) and denominator (D) of the objective function will be referred to as **AS-SPoC**.

Table 4.1: **Overview of introduced SPoC regularization variants.** Two model selection schemes are applied: cross-validation (CV) and based on the Ledoit & Wolf shrinkage estimator (LW). For $\widetilde{\mathbf{\Sigma}}(e)$ it is explicitly marked if regularization is applied to the numerator (N) and/or to the denominator (D) of the objective function. The checkmarks refer to covariance normalization.

Method	Regularization Target		Normalization
	$\widetilde{\mathbf{\Sigma}}(e)$	$\widetilde{\mathbf{\Sigma}}_{avg}$	$\widehat{\mathbf{\Sigma}}(e)$, $\widehat{\mathbf{\Sigma}}_{avg}$
Tik-SPoC	-	CV	-
NTik-SPoC	-	CV	✓
ASNTik-SPoC	LW: N	CV	✓
AS-SPoC	LW: N,D	-	-
aTik-SPoC	-	LW	-

4.3 EXPERIMENTS AND VALIDATION PROCEDURE

4.3.1 *Simulation data scenario*

In this work, we aim to characterize and benchmark the introduced regularization techniques to the SPoC$_\lambda$ algorithm. However, in the majority of real-world EEG experiments there is no ground truth source available, which severely challenges the validation procedure. To compensate for this, a novel data-driven simulation approach for labeling datasets was utilized [189]. It generates ground-truth label information based on known sources from arbitrary pre-recorded EEG measurements. This post-hoc data

labeling allowed obtaining noiseless labels from a relatively large amount of EEG data (here up to 1000 epochs) while conserving the real statistics of the neural activity including non-stationarities of the signal. Furthermore, the approach provided full control over label noise and allows studying its influence upon the decoding performance. In the following, a detailed description for the dataset generation is given.

Table 4.2: **Overview of two additional datasets used for validating the regularization methods.** Each dataset contains one offline session of each subject.

	dataset (D2)	**dataset (D3)**
No. of subjects (after preprocessing)	40 (12)	11 (7)
Experimental setup		
Paradigm	motor imagery	steady-state stimulation
Recorded EEG channels N_{ch} (utilized)	118 (63)	63 (63)
Reference publication	[190]	[94]

4.3.1.1 *Dataset and preprocessing*

As listed in Tab. 4.2, the dataset (D2) consists of 40 single-session motor imagery recordings which formed the basis for the simulation. The experimental design of the motor imagery paradigm is described in detail in [190]. From the recorded EEG, we utilized the signals of 63 passive EEG channels placed according to the extended 10-20 system. The preprocessing of each raw EEG dataset consisted of a low-pass filtering at 48 Hz, a sub-sampling to 120 Hz and a final high-pass filtering at 0.2 Hz. For each dataset, the continuous EEG recordings of active task periods (from the task cue to the end of the imagery interval) were segmented into non-overlapping epochs of 1 s duration. Artifact epochs were identified by a min-max threshold and by a variance criterion. The latter was additionally applied to detect and remove outlier channels. Details about the artifact preprocessing are described in Sec. 3.2.4.

As the robustness of regularized SPoC variants depending on the number of training epochs N_e shall be studied here, we discarded datasets with $N_e < 1000$. Similarly, datasets in which more than 10% of the original EEG channels had to be rejected, have been removed from further analysis. Applying these rather strict criteria, the data of 12 out of 40 subjects remained. While a relaxation of the preprocessing criteria would have allowed to include more subjects, we have decided for evaluating the methods on large, well-preprocessed EEG datasets that provide a comparable dimensionality across subjects.

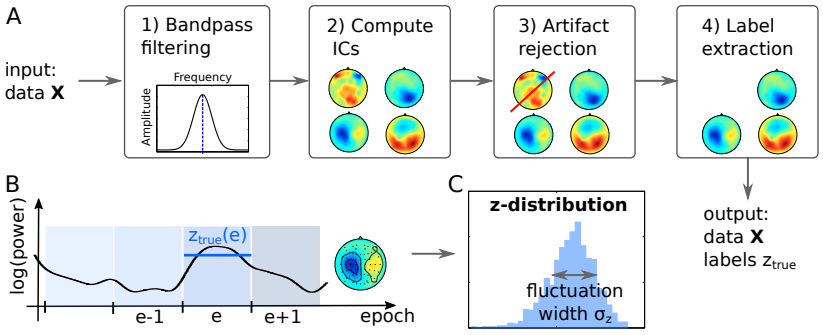


Figure 4.1: **Data-driven post-hoc labeling of arbitrary pre-recorded EEG signals.** (A) Processing pipeline to extract independent components (ICs). (B) For each IC and epoch e, the log-bandpower average of the epoch serves as a ground truth label $z_{true}(e)$. (C) Distribution of z_{true} over all epochs of an exemplary IC. Its bandpower fluctuation width is described by σ_z.

4.3.1.2 *Post-hoc labeling of pre-recorded EEG data*

As illustrated in Fig. 4.1 (A), the following steps were applied to generate continuous labels z_{true} from pre-recorded EEG datasets in a data-driven way:

1. Bandpass filtering of the data to a frequency band of interest. For our analysis, we choose the alpha-band frequency range of $[8, 12]$ Hz.

2. Based on the bandpass filtered data, an ICA decomposition (fastICA, [85]) into $N_{in} = 20$ independent components (ICs) was computed.

3. To identify and remove artifactual components in an automatic way, the data-driven classification approach MARA [89] for the identification of artifactual components was applied. A posterior probability threshold ($p_{art} = 10^{-8}$) describing the probability of an artifact feature was applied for discarding components of non-neural origin resulting in $N_{sel} \leq N_{in}$ selected ICs.

4. The log-bandpower for each selected component j with $j = 1, .., N_{sel}$ was computed by the Hilbert transform and averaged in each $1\,s$ time interval which defined the epoch-wise known target variable $z_j(e)$ as sketched in Fig. 4.1 (B).

Overall, the preprocessed data of 12 subjects resulted in 145 oscillatory components ($\approx$ 12 per subject) which survived MARA. For each selected IC, the log-bandpower activation was sampled across $N_e = 1000$ epochs and thus delivered continuous epoch-wise labels $z_{true}(e)$ for the respective epoched EEG signals $\mathbf{X}(e)$.

We expect the SPoC decoding accuracy to be sensitive to the strength of envelope changes of an oscillatory component. The simulation design enables to empirically study this influence by extracting the absolute width in bandpower fluctuations of a single selected IC across a full session.

Therefore, we define the fluctuation width of the jth IC as $\sigma_z := \mathrm{Var}[z_j(e)]$ calculated across the $N_e = 1000$ epochs as illustrated in Fig. 4.1 (C).

4.3.1.3 *Probing the algorithms under reduced datasets and label noise*

In an offline analysis using the generated 145 labeled datasets, all five introduced SPoC regularization variants and the standard SPoC approach were evaluated in a 10-fold chronological CV. For each epoch e, an estimate of the target variable $z_{est}(e)$ was derived according to Eq. (3.1) by applying the highest ranked spatial filters obtained from the training data.

To analyze the benefit of regularization under different dataset sizes, we evaluated the algorithms' stability by systematically reducing each dataset with originally 1000 epochs to smaller data chunks. Therefore, epochs from the session end were removed. For each of the 145 labeled datasets, 22 discrete, logarithmically scaled dataset sizes $N_e \in [20, 1000]$ (respectively training set sizes N_{train}) were tested. Similarly, we probed the stability of our approaches under varying label noise conditions. Therefore, each sample of the target variable distribution z_{true} was modified by adding normally distributed label noise, resulting in a noisy label set z_{noisy} which was used for the CV procedure. According to the label noise model proposed by Castaño-Candamil et al. [189], the correlation between the undistorted and the noisy labels $\rho_n = \mathrm{Corr}(z_{true}, z_{noisy})$ can be controlled via the label noise parameter $\xi_n := 1 - \rho_n$. A value $\xi_n = 1$ refers to maximal label noise, while $\xi_n = 0$ indicates that the labels are completely noise free. Five fixed levels for ξ_n were evaluated.

For the CV-based regularized SPoC variants (see Tab. 4.1), the regularization strength α was varied in a range $\alpha \in \{0; [10^{-8}, 10^{0}]\}$. Overall, 20 discrete, logarithmically scaled α levels were analyzed. To summarize, we tested all algorithms on different hyperparameter configurations $\omega \in \{(N_{train}, \xi_n, \alpha)\}$.

4.3.2 *Real-world data scenario*

4.3.2.1 *Datasets for evaluation*

For the examination of the regularization methods in real-world decoding scenarios, we utilized two different datasets, which will shortly be described hereafter including the applied preprocessing steps.

Dataset (D3): Decoding auditory stimulus intensity

Tab. 4.2 contains dataset (D3) with experimental data on steady-state auditory evoked potentials (SSAEP) conducted in a single session with 11 subjects. These data were used to validate the original non-regularized SPoC algorithm [94]. The experiment builds upon a known gain of EEG bandpower with increased stimulus intensity [191]. Therefore, the stimuli were generated with a carrier frequency of 500 Hz and an additional modulation by a sinusoidal 40 Hz signal which resulted in a so-called

steady-state stimulation. The intensity of the stimuli were systematically varied in the range of 10 to 35 dB. A single session consisted of 3 blocks of 5 min ongoing stimulation. In total, 7 out of 11 subjects provided the required SSAEP features and were used for further evaluation.

Throughout the experiment, EEG data was sampled with 1 kHz from 63 passive Ag/AgCl electrodes positioned according to the 10-20 system. All channels were referenced to the nose. For the offline analysis, signals were low-pass filtered with a cut-off at 90 Hz, a notch filter applied around 50 Hz and the data downsampled to 250 Hz. Finally, the EEG data were bandpass filtered to a frequency range of $[39, 41]$ Hz (centered around the 40 Hz modulation frequency) and epoched into segmented data $\mathbf{X}$ of 1 s length. Overall, data of each subject provided at least $N_e = 897$ epochs. The epoch-wise target variable $z(e)$ was computed as the squared stimulus intensity in dB. For the evaluation of the regularization variants, which are expected to be beneficial in small-dataset scenarios, the data $\mathbf{X}$ and z of each subject were split in three disjunct datasets of size $N_e = \{125, 250, 500\}$ leading to 21 datasets for evaluation.

Datasets (D1a) and (D1b): Decoding single-trial motor performance

Referring to the datasets (D1a) and (D1b) described in Sec. 3.2.2, we utilized single-session data of 25 subjects who participated in the SVIPT paradigm with 400 or 240 trials per session. The task enabled to extract a single-trial motor performance metric such as reaction time (RT). As a major finding of the previous Chap. 3, we reported that oscillatory bandpower features recorded during the pre-go interval can partially explain upcoming single-trial motor performance.

Building upon these findings, here the EEG signals were segmented into epochs along the time interval $[-500, +500]$ ms relative to the go-cue in each trial to decode RT of the upcoming motor task. By this choice of the time interval, we do not predict upcoming performance any more but therefore have potentially larger absolute decoding performance levels. After data preprocessing and outlier rejection as described in Sec. 3.2.4, we now restricted any further analysis to oscillatory features within the alpha-band frequency range of $[8, 12]$ Hz. The bandpass filter was realized applying a zero-phase Butterworth filter of 6^{th} order. Overall, data of 24 subjects were used for further analysis which provided between 186 and 360 epochs after preprocessing (at least 150 epochs were required).

4.3.2.2 *Evaluation scheme*

All algorithms were evaluated within a (nested) 10-fold chronological CV. The three CV-based regularization variants demanded an additional inner CV to estimate the individually optimal regularization parameter α^*. It was chosen among 15 discrete, logarithmically scaled values in the range $\alpha \in [10^{-8}, 1]$. The α-value maximizing the z-AUC evaluation score (details see Sec. 4.3.3) was selected and applied to the outer CV in order to train the respective spatial filtering algorithm and linear regression model. The

methods aTik-SPoC and AS-SPoC allow for an analytical estimate of α^* and hence did not require an inner CV.

For each α in the inner or outer CV, the following scheme was applied: a spatial filter set $\{\mathbf{w}^{(i)}\}_{i=1,..,N_c}$ was computed on training data $\mathbf{X}_{tr}$. The first $N_{feat} = 4$ highest ranked components were selected as input to train a linear regression model with coefficients $\{\beta_j\}_{j=0,..,N_{feat}}$. The model was trained upon the log-bandpower features $\Phi_{j,tr} = \log(\mathrm{Var}[\mathbf{w}_{tr}^{(j)}\mathbf{X}_{tr}])$. On each feature $\Phi_{j,tr}$, the mean $\mu_{j,tr}$ and the variance $\sigma_{j,tr}$ was estimated in order to standardize the data to zero mean and unit variance before entering the regression model. Given unseen test data $\mathbf{X}_{te}$, the log-bandpower features $\Phi_{j,te}(e) = \log(\mathrm{Var}[\mathbf{w}_{tr}^{(j)}\mathbf{X}_{te}])(e)$ for each selected spatial filter $\mathbf{w}_{tr}^{(j)}$ were first standardized by $\mu_{j,tr}$ and $\sigma_{j,tr}$. Subsequently, the corresponding coefficients β_j of the trained linear regression model enabled to estimate the target variable $z_{est}(e)$ via:

$$z_{est}(e) = \beta_0 + \sum_{j=1}^{N_{feat}} \beta_j \Phi_{j,te}(e) \tag{4.7}$$

In contrast to the simulation scenario, the total number of ground truth neural source(s) which might (partially) explain the target variable z_{true} is not known a priori. Thus by applying a regression model, we assume that several sources might contribute to explain the labels z_{true} of the real-world decoding scenarios.

4.3.3 *Evaluation scores*

To compare the estimated labels z_{est} with the known or measured labels z_{true} in the simulation and real-world scenarios across the proposed regularization variants, different evaluation scores can be considered [28]. In general, the Pearson correlation coefficient could be utilized but has the drawback, that it is very sensitive to the number of samples [192]. Therefore, we instead decided to utilize the following three scores in this chapter:

- **Angle θ between spatial filters**: The design of the simulation scenario gives access to each ground truth spatial filter $\mathbf{w}_{true}$. As all proposed SPoC variants directly optimize for a spatial filter estimate $\mathbf{w}$ with arbitrary sign and amplitude (this characteristic is inherited from the formulation as an eigenvalue problem), the angle θ between the spatial filters can directly serve as an evaluation metric:

$$\theta_r = \arccos\left(\frac{\mathbf{w}^\top \mathbf{w}_{true}}{\|\mathbf{w}\| \, \|\mathbf{w}_{true}\|}\right)$$

$$\theta = \begin{cases} \theta_r, & \theta_r \leq \pi/2 \\ \pi - \theta_r, & \theta_r > \pi/2 \end{cases}$$

  with $0 < \theta < \pi$. A perfect decoding will be expressed by an angle $\theta = 0$. Please note that the angle θ can only be estimated within the simulation scenario.

- **Separability z-AUC of labels**: As introduced in Sec. 3.2.7, the continuous labels z_{true} can be transferred into a two-class scenario according to the median of z_{true}. Then the area under the ROC curve describes the separability of the labels, named z-AUC.

- **Relative z-AUC performance**: The score z-AUC$_{ref}$ corresponds to the baseline performance of SPoC without any regularization. Here, we will compare it to performances obtained by the proposed regularized variants (see Tab. 4.1). Given a hyperparameter configuration ω, the target variable obtained under these hyperparameters $z_{est}(\mathbf{w}(\omega))$ can be estimated using Eq. (3.1) and the corresponding z-AUC can be computed. For fixed ω, the performance of a regularized SPoC variant z-AUC$_{reg}(\omega)$ can be assessed as the relative change of z-AUC to the baseline SPoC performance:

$$\text{rel. z-AUC}(\omega) := \frac{\text{z-AUC}_{reg}(\omega) - \text{z-AUC}_{ref}(\omega)}{\text{z-AUC}_{ref}(\omega)} \tag{4.8}$$

  If rel. z-AUC > 0, this directly corresponds to a relative performance increase compared to standard SPoC and vice versa.

4.4 RESULTS

First, we studied the characteristics of the regularization algorithms on 145 analysis problems within the simulation framework. It allows assessing the influence of (hyper)parameters such as regularization strength, dataset size or label noise under controlled conditions. Second, the approaches were tested on two real-world datasets to verify the transferability of the findings and to provide rules of thumb for the practitioner.

4.4.1 *Simulation data*

4.4.1.1 *Component labeling according to bandpower fluctuation width*

The SPoC algorithm optimizes for oscillatory components that co-modulate in their bandpower with a given target variable. In Fig. 4.2, the relation between the fluctuation width σ_z and the baseline SPoC performance z-AUC$_{\text{ref}}$ on the full dataset $N_e = 1000$ is shown for each of the 145 ICs (correlation $R = 0.31$ with $p = 2.20 \cdot 10^{-4}$). The results indicate that the decoding quality of SPoC depends on the fluctuation width σ_z of the underlying neural component, with stronger fluctuation width being related to higher decoding quality. For further analysis of the simulation data, all 145 ICs were labeled according to their bandpower fluctuation width σ_z into three classes determined by the lower and upper quartile according to the distribution of σ_z across all components (see color coding in Fig. 4.2). In the following, we will show the decoding performances z-AUC$_{\text{GA}}$ and θ_{GA} as grand average for each corresponding fluctuation width class.

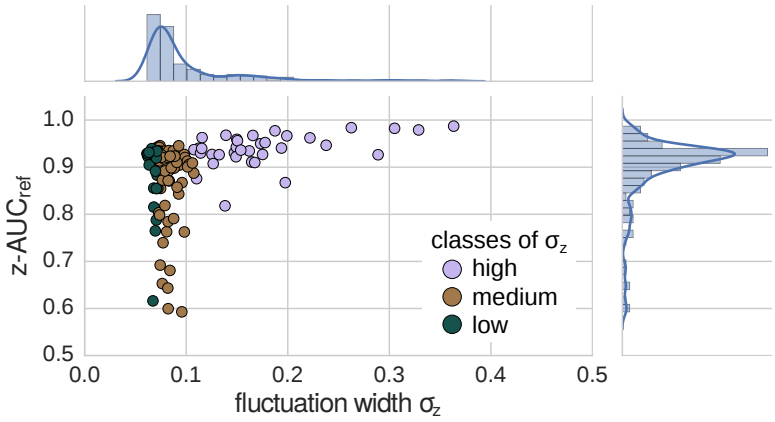


Figure 4.2: **Simulation data: component labeling according to fluctuation width.** Scatter plot relating the fluctuation width σ_z of each selected ICA component to their baseline SPoC performance z-AUC$_{ref}$ for non-reduced datasets with 1000 epochs. Based on its σ_z-distribution, the dataset of each IC was labeled into one of three classes, defined by the quartile thresholds Q_{25} and Q_{75}.

4.4.1.2 *Sensitivity to regularization parameter*

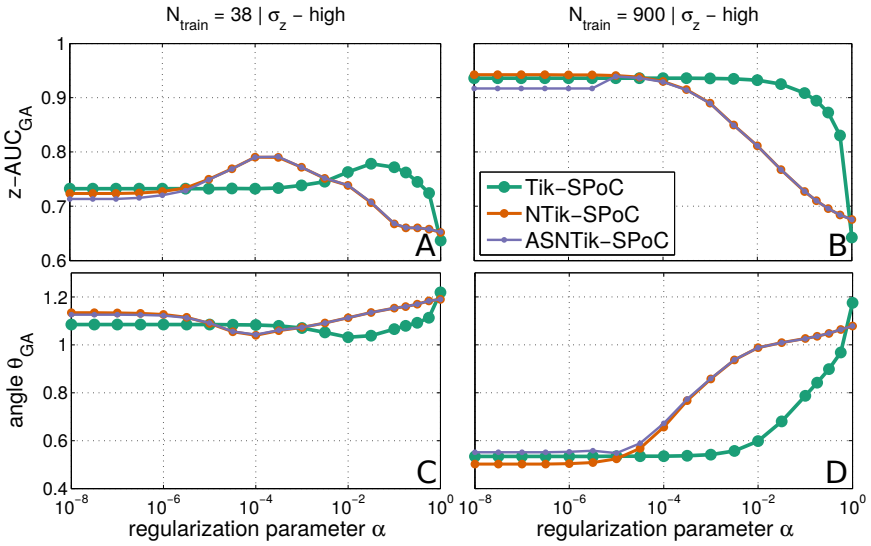


Figure 4.3: **Simulation results: influence of regularization strength α onto the decoding accuracy of three SPoC variants regularized via CV.** The grand average performance z-AUC$_{GA}$ is reported in the top row (subplots (A) and (B)), while subplots (C) and (D) in the lower row report the angle θ_{GA} between the estimated highest ranked and the ground truth filter as evaluation score. Subplots in the left and right columns differ in the number of training data points (epochs) used for SPoC decoding. Results are reported for the class 'σ_z - high'.

Regarding the CV-based regularized SPoC versions, their sensitivity to the regularization parameter α is reported in Fig. 4.3 exemplarily for the 'high σ_z' class. It reflects the grand average (GA) of all components contained in this class and provides different evaluation scores. The first

row reports the z-AUC$_{GA}$ while the second row summarizes the angle θ_{GA} between filters. A regularization benefit is expressed via an increasing z-AUC$_{GA}$ or a decreasing θ_{GA} relative to the performance level at $\alpha = 10^{-8}$. A few observations can be summarized from Fig. 4.3: First, the two evaluation scores z-AUC and θ are highly (anti-)correlated across the shown dataset scenarios and SPoC regularization variants. As in real-world data the ground truth will not be known a priori, further analysis will need to be restricted to the metric z-AUC. Second, an increase of the training set size N_{train} (left to right column) leads to a lower sensitivity wrt. α. Third, a comparison of α sensitivity ranges across the three regularization variants yields that NTik-SPoC and ASNTik-SPoC are sensitive in the interval $10^{-6} \leq \alpha \leq 1$ while Tik-SPoC is only sensitive within $10^{-3} \leq \alpha \leq 1$. Fourth, NTik-SPoC and ASNTik-SPoC behave highly similar, while Tik-SPoC shows a qualitatively different behavior. Based on these observations, further analysis will focus on differences between NTik-SPoC and Tik-SPoC. Fifth, extreme regularization with $\alpha = 1$ leads to a drop of decoding performance regardless of the approach, while in the absence of regularization ($\alpha = 0$) a slight improvement due to trace normalization can be reported for NTik-SPoC.

4.4.1.3 *Influence of reduced datasets and fluctuation width*

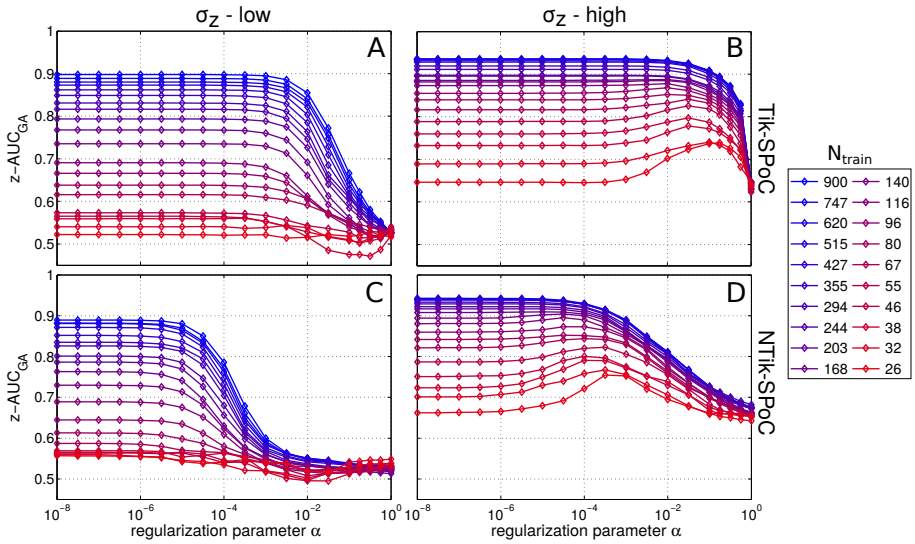


Figure 4.4: **Simulation results: sensitivity of regularized SPoC variants to α and to reduced training set sizes N_{train}.** The grand average performance z-AUC$_{GA}$ is reported for Tik-SPoC (top row) and NTik-SPoC (bottom row) and separately for the fluctuation width classes 'low' (left column) and 'high' (right column).

The simulation scenario grants access to test the stability of different regularized SPoC variants under reduced datasets. For the CV-based methods, a sensitivity analysis for the regularization strength α under 22 training set sizes N_{train} is shown in Fig. 4.4 for Tik-SPoC (first row) and NTik-SPoC (second row). The two columns in Fig. 4.4 reveal the influence of the com-

ponents' fluctuation width σ_z (left: low, right: high). We observed, that regularization has the strongest effects for components with large σ_z and for small training sets. With increasing training set size N_{train}, the sensitivity range for α shifts towards smaller α values. Comparing the depicted methods, NTik-SPoC shows a higher sensitivity to regularization strength α compared to Tik-SPoC. Interestingly, for all subplots (A)-(D) the curves along different N_{train} values converge at $\alpha = 1$, as for this value the SPoC methods collapse to a PCA on the z-weighted covariance. Even for this extreme choice of α, data characterized by higher σ_z reaches a better decoding performance than data with lower σ_z.

To quantify the decoding performances across methods, the maximum GA performance z-AUC$_{\max} :=$ z-AUC$_{\mathrm{GA}}(\alpha^*)$ is reported in Fig.4.5 (A) and (B) in the absence of label noise. Therefore, the optimal regularization strength $\alpha^* = \arg\max_\alpha$ z-AUC$_{GA}(\alpha)$ is selected for fixed N_{train} and σ_z class. For variants using the LW estimate, this selection is not necessary as there is an analytic solution for α^* such that z-AUC$_{\max}$ = z-AUC$_{\mathrm{GA}}$. Accordingly, the relative performance change rel. z-AUC(α^*) is reported on the GA level in Fig. 4.5 (C) and (D), while (E) and (F) report the statistical significance of the findings. Therefore, a one-sided Wilcoxon signed-rank test was applied to test if the median of performance differences (z-AUC$_{max,ref}(\omega)$ − z-AUC$_{max,reg}(\omega)$) is smaller or equal to zero for fixed N_{train} and σ_z. If a p-value $p < 0.05$ was found (not corrected for multiple testing), the configuration ω reveals a significant difference among the two methods, indicated by a colored data point in Fig. 4.5 (E) and (F). The following observations can be reported: First, the absolute decoding performance strongly depends on N_{train} regardless of the regularization method and σ_z class. Second, there is a relative performance increase of all introduced regularization methods up to training sets of size $N_{train} \approx 60$ on the grand average level. For larger datasets the regularization does not reveal an additional benefit on the grand average. Third, our results indicate that regularization is beneficial for various methods in the 'high σ_z' class, while this is not the case for 'low σ_z'. Here, a noticeable case is reported by the performance of AS-SPoC which drastically loses performance for $N_{train} \gtrsim 50$.

4.4.1.4 *Stability under label noise and reduced datasets*

As in most real-world scenarios label noise challenges the decoding performance of subspace methods like SPoC [31]. Thus, we studied its influence for reduced datasets within the simulation data. Fig. 4.6 exemplary shows the degrading decoding performance under label noise conditions for aTik-SPoC and AS-SPoC for 'high σ_z'. Both methods have in common, that performance estimates are very noisy under small dataset size and increasing label noise. Regarding the maximally achievable decoding performance for both methods at $N_{train} = 900$, the absolute performance z-AUC$_{\mathrm{GA}}$ scales almost linearly with the amount of label noise ξ_n. Referring to the relative performance change shown in (C) and (D) as well as the statistical tests in (E) and (F), they reveal that under increased levels of label noise ξ_n

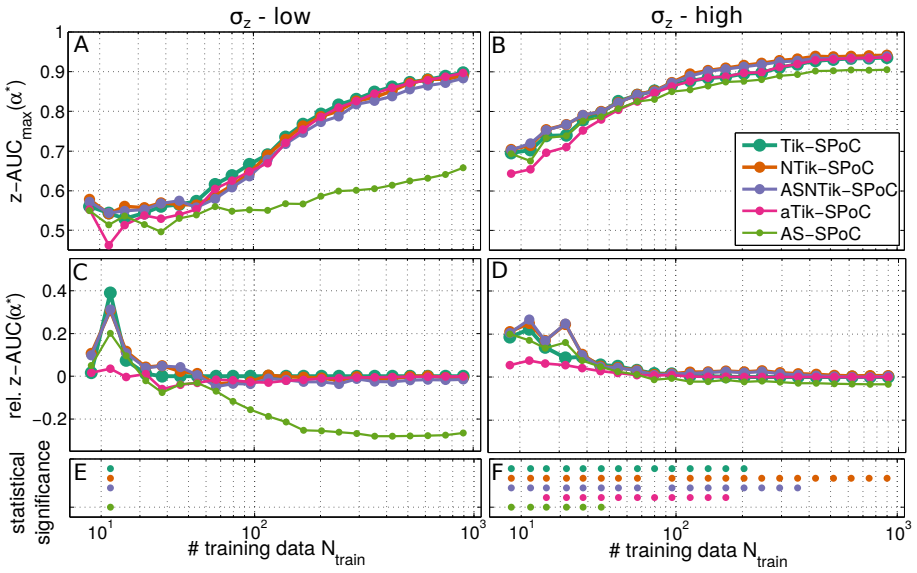


Figure 4.5: **Simulation results: influence of training set size and fluctuation width upon decoding performance of optimal regularization strength α^*.** The top row depicts the grand average **absolute** performance of five regularized SPoC variants for ICs that either have low (A) or high (B) bandpower fluctuation width. The middle row depicts performance increase or decrease of the five regularized methods **relative** to the baseline SPoC method without any regularization and again separately for IC's of low (C) and high (D) fluctuation width. Subplots (E) and (F) reveal color-coded points for each training set size in which the regularized variant significantly outperformed the baseline method (Wilcoxon signed-rank test with $p < 0.05$).

even larger training sets can profit from regularization when compared to the unregularized SPoC. While for $\xi_n = 0$ a relative performance increase on the GA can be found up to $N_{train} \approx 60$, for $\xi_n = 0.6$ it increases up to $N_{train} \approx 800$. This effect is stronger for AS-SPoC than for aTik-SPoC. Despite not shown here, we would like to mention, that under increased label noise the performance gain of the regularized variants with larger N_{train} can be observed also for the 'low σ_z' case, but with a lower overall decoding performance.

4.4.1.5 *Optimal regularization parameter ranges*

To identify suitable ranges of the regularization parameter for the CV-based methods, color-coded contour maps of relative performance changes are provided in Fig. 4.7. The maps show the grand average rel. z-AUC$_{GA}$ within the (N_{train}, α) hyperparameter space separately for the two methods Tik-SPoC (first column) and NTik-SPoC (second column). Maps in the upper row summarize the performance changes in the absence of label noise ($\xi_n = 0$) while the lower one provides these results under systematic label noise ($\xi_n = 0.4$). The blue areas in each map mark ranges in the hyperparameter space, in which a relative performance increase is obtained, while "no-go" areas in red associate with a decrease of decoding quality. When comparing

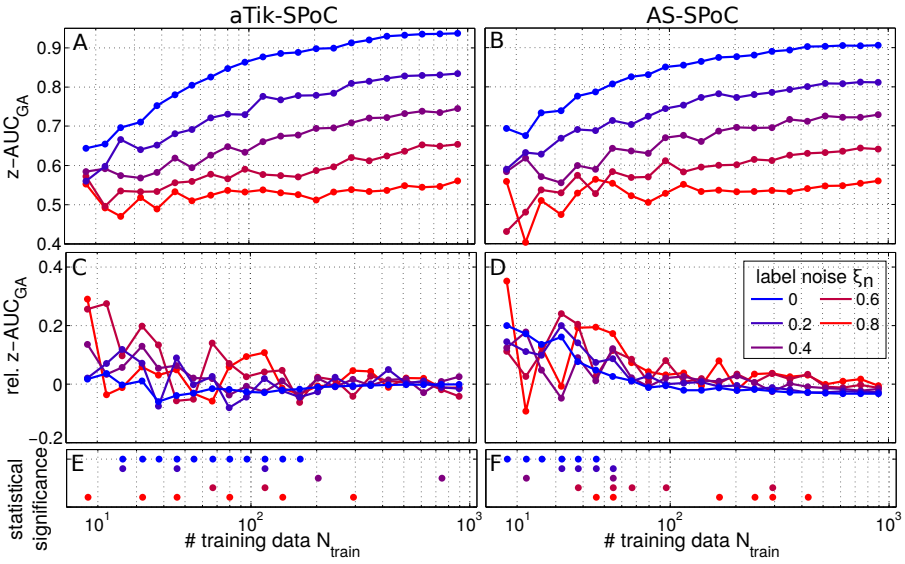


Figure 4.6: **Simulation results: interaction between label noise level ξ_n and dataset size N_{train}.** A level of $\xi_n = 0$ states the absence of label noise. All curves report the grand average results for ICs belonging to the 'high σ_z' class. Subplots (A) and (B) in the top row provide the **absolute** grand average performances for aTik-SPoC and AS-SPoC, while the middle row depicts **relative** performance changes. The dots in (E) and (F) indicate configurations, for which the regularized variant significantly outperformed the baseline method (Wilcoxon signed-rank test with $p < 0.05$).

Fig. 4.7 (A) and (B), we observe that the trace norm in NTik-SPoC induces a reduction of optimal α values by a few orders of magnitude as well as a larger sensitivity range compared to Tik-SPoC. Both plots reveal consistently a "no-go" area towards the top right corner, which indicates, that on the grand average strong regularization is detrimental, when large training datasets without label noise are available. With additional label noise in Fig. 4.7 (C) and (D), the heterogeneity of the relative performance landscape increases and the "no-go" areas at the top right shift towards larger N_{train}. In accordance with the automatic shrinkage based methods visualized in Fig. 4.6 we find, that the inclusion of label noise ξ_n into the simulation has the effect that regularization might even be beneficial for large training sets.

For different training set sizes N_{train}, we now compare the CV-based estimates of α with those of aTik-SPoC, which makes use of an analytical solution α^*. The grand average of α^* is plotted into Fig. 4.7 (B) and (D). As the analytical solution for α^* [186, 187] is proportional to N_{train}^{-2}, it should scale anti-proportional with $\log_{10}(N_{train})$, which in fact was observed in Fig. 4.7 (B). It is worth mentioning, that the analytic choices of α^* are not influenced by label noise – compare maps (B) and (D) – as the involved covariance shrinkage (see Eq. (4.6)) does not make use of the label information.

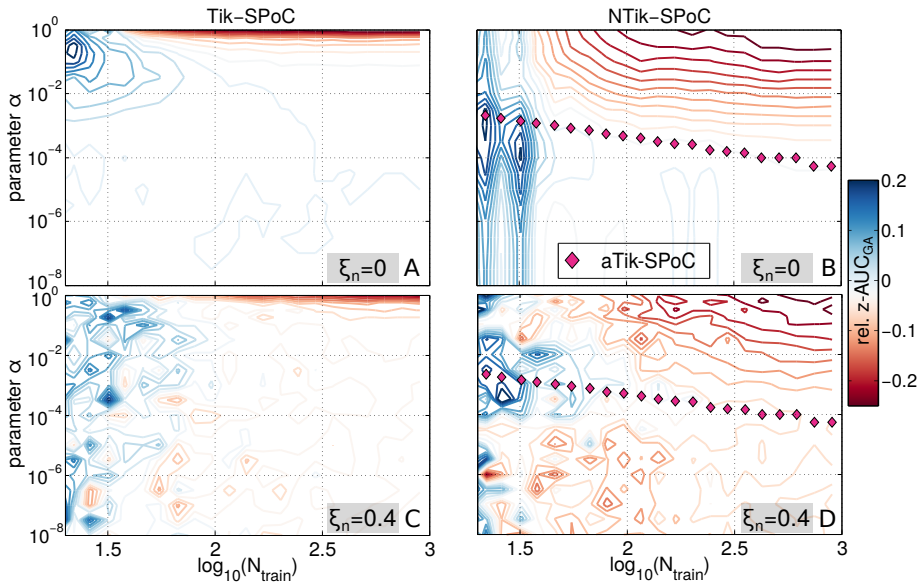


Figure 4.7: **Simulation results: landscape of the grand average relative performance changes.** The z-AUC is shown dependent on the training set size N_{train} and regularization strength α for ICs of high fluctuation width σ_z. The isolines of relative performance changes were interpolated along a grid search. No label noise was applied to generate maps (A) and (B) for methods Tik-SPoC and NTik-SPoC, respectively. The second row reports the landscapes including a label noise level of $\xi_n = 0.4$ for both methods. Additional diamond markers in subplots (B) and (D) depict the grand average of α^* for aTik-SPoC, which is independent of label noise. This method utilizes analytically derived values of α^* and may serve as a reference for the CV-based NTik-SPoC.

4.4.2 *Real-world data*

4.4.2.1 *Comparison of regularized SPoC variants*

Fig. 4.8 contains the subject-wise performance comparison of all regularized SPoC variants to standard SPoC for the two investigated datasets. To compare each regularized variant to its baseline, we report the overall ratio of subjects for which the regularized variant outperforms standard SPoC. To verify if a regularized variant reaches a statistically significantly higher performance compared to standard SPoC, a two-sided Wilcoxon signed-rank test was evaluated on the group level. If the p-value was below 0.01, the top left subplot corner is annotated by a star symbol.

The following observations can be stated on the two real-world datasets: The performance changes induced by aTik-SPoC are negligible small regardless of the underlying dataset and in contrast to all other regularization variants. Across the remaining approaches, the CV-based NTik-SPoC and ASNTik-SPoC variants behave very similarly, which has been observed before on simulation data (see Fig. 4.5).

Regarding the results on the SSAEP dataset (D_3) in Fig. 4.8 (A)-(D), all three CV-based variants as well as AS-SPoC significantly outperform the

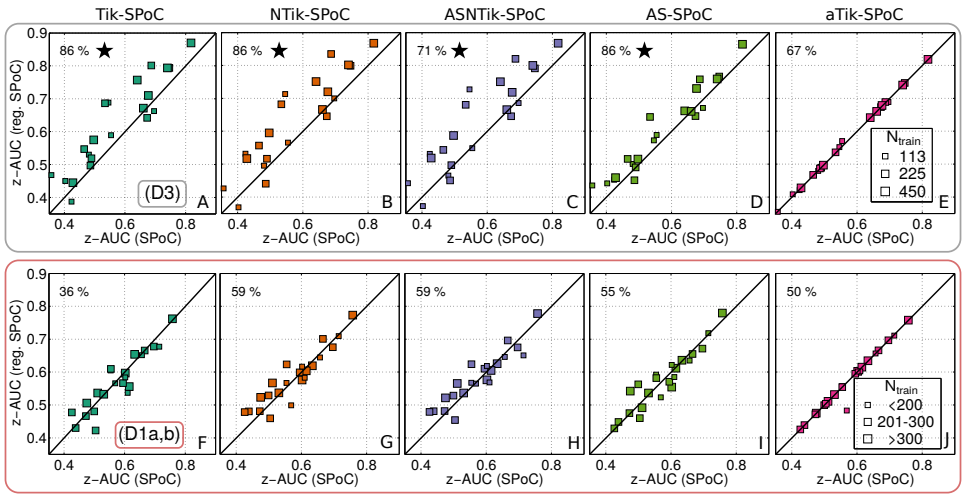


Figure 4.8: **Real data: performance of different regularized SPoC variants for two datasets.** Single scatter plots reveal a comparison with the unregularized baseline method SPoC. Each column refers to a regularization variant. The first row reports results for dataset (D3), the second row for (D1a,b). In each scatter plot, the marker size encodes the underlying training dataset size. At the top left, the percentage of points for which the regularized variant outperforms baseline SPoC is given. A star symbol reports a significance level of $p < 0.01$.

baseline SPoC variant. Moreover, a regularization benefit was found for both, initially poor and well performing models. This finding, however, does not transfer to the results achieved for dataset (D1a,b), as shown in Fig. 4.8 (F)-(J). Here, we can state a tendency towards larger benefits for initially poorly performing subjects. On the group level, all regularization methods except Tik-SPoC registered the majority of data points above the bisectrix, while none of the approaches significantly outperform the baseline SPoC performance.

4.4.2.2 *Selected regularization strengths*

The regularization parameter α_f obtained on real-world data by the nested CV-based regularization variants across folds f are reported in the first two columns of Fig. 4.9. The plots should be compared with the maps for simulations addressed in Fig. 4.7. The median regularization strength across folds is plotted against the training set size and color coded by the associated z-AUC performance. The results for both real-world datasets indicate that NTik-SPoC operates in smaller absolute α ranges than Tik-SPoC does, which is in accordance with the observations from the simulation in Fig. 4.7. For both datasets, the regularization strength is in almost all cases outside the "no-go" areas of the simulation as α was selected by nested CV from the interval $[10^{-8}, 1]$. For only a few subjects of dataset (D1a,b), a large α was chosen for Tik-SPoC. As expected from simulations, this strong regularization is linked with a low absolute decoding level. Interestingly, comparing α-levels for NTik-SPoC across both datasets, it reveals mostly stronger regularization for the SSAEP data than for the SVIPT scenario.

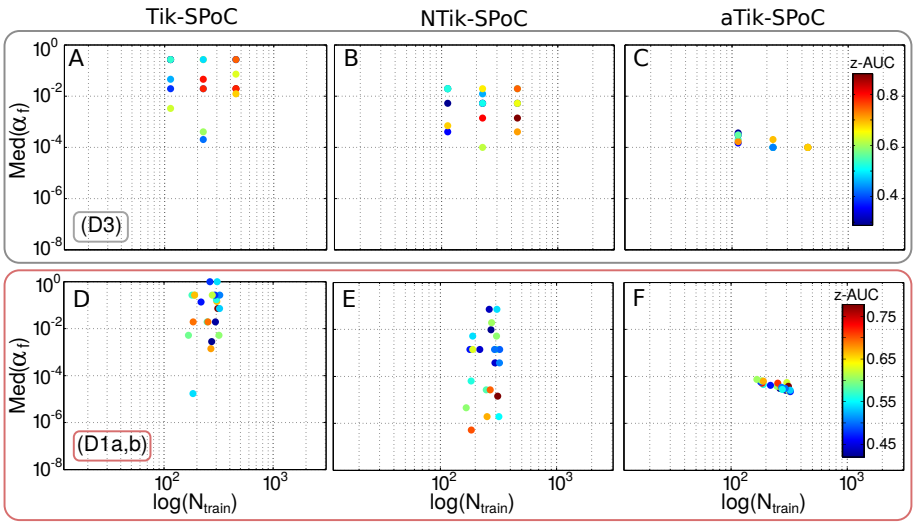


Figure 4.9: **Real data: averaged regularization strength for three different methods.** As an average estimator, the median regularization strength is computed across the 10-fold chronological cross-validation for each dataset depending on the training set size N_{train} and color coded by the achieved z-AUC decoding performance. The first row reports results for dataset (D3), the second row for (D1a,b). Each column refers to a regularized SPoC variant.

The median of the analytically computed α_f^* across folds for aTik-SPoC is presented in Fig. 4.9 (C) and (F). On dataset (D3), allmost all and on (D1a,b) most subjects reveal a way smaller median regularization strength compared to the CV-based NTik-SPoC method. As previously reported in Fig. 4.8 (D) and (I), the analytical solution does not elicit a significant decoding improvement.

To gain an intuition about the effect of the different regularization strengths on resulting spatial activity patterns, an example regularization parameter sweep is reported for an exemplary subject in the appendix A.2.

4.5 DISCUSSION

In summary, we have proposed a set of novel regularization techniques for SPoC. We investigated their effectiveness by evaluating their performance both on simulated and on real-world datasets. Overall, NTik-SPoC based on Tikhonov regularization and additional covariance normalization turned out to be the most beneficial technique.

4.5.1 *Simulation scenario*

A closer look upon the simulation results clearly shows that the regularization benefit for SPoC strongly depends on the dataset size, prevalent label noise conditions as well as on the fluctuation width of the underlying component. As a strong absolute performance variability across datasets was present in the simulation, the reported grand average performance

provides a way less optimistic view than single dataset results do. The largest regularization benefit was reported for low amount of data and components with large fluctuation widths. The latter observation might be explained by the intrinsic difficulty of SPoC to recover sources of small bandpower changes.

Intuitively, additional label noise reduces the information content per data point such that the estimation of $\mathbf{\Sigma}_z$ gets more demanding. Theoretically, this disadvantage could be compensated by either enlarging the training set or by adding regularization. Using the large amount of simulation data, we were able to show that under label noise conditions even larger datasets profit from regularization.

Surprisingly, in the simulation we found that AS-SPoC loses performance for large datasets (especially for 'low σ_z' components) while it outperformed standard SPoC on small datasets and revealed a good performance on real-world data as well. This observation might be explained as follows: In the simulation data, the target variable is directly estimated from the EEG (IC) epoch. As such, there should be enough samples in each epoch to estimate reliably the target variable, since it was created this way. Epoch regularization might thus not be necessary here. However, for real data this might not be the case, as the target variable is not directly dependent on the EEG epoch and contains an even unknown label noise level. As such, epoch regularization might be much more useful in that case.

The direct transferability of the simulation results to real-world data is limited by three major differences: First, in real-world experiments the number of neural sources is not known a priori. Thus, a good decoding of source power typically requires the use of several components and of a regression model. Second, in real-world experiments both, label noise and the components' fluctuation widths act as latent variables and cannot directly be estimated. Third, while in the simulation we can almost perfectly recover the label information given sufficient amount of data (z-AUC > 0.9), in real-world experiments we clearly expect a decreased upper limit of the decoding performance—as solely bandpower information may not suffice to fully explain the labels—which even strongly depends on the experimental setting.

4.5.2 *Real-world scenarios*

The two utilized real-world datasets (D1a,b) and (D3) showed slightly different outcomes in terms of the tested regularization variants. For the SSAEP dataset a regularization benefit for the CV-based variants and AS-SPoC was found irrespectively of baseline performance. On single datasets, improvements of the decoding accuracy of up to 27 % was found. The more challenging SVIPT data (D1a,b) revealed that predominantly models of initially poorly performing subjects were improved by the aforementioned regularized SPoC approaches. These different effects might be explained by varying characteristics of the used datasets, such as the non-comparable upper limits and ranges of achievable decoding accuracies as well as

prevalent label noise conditions. An indicator for a large regularization benefit might be the CV-based selected regularization strength, as we observed that the SSAEP data required larger α values than the SVIPT data (see Fig. 4.9). So far, we cannot report a single regularization variant that systematically performed best on all tested datasets. Following the recently proposed large-scale benchmarking concept [193], a future evaluation on a multitude of real-world regression datasets—which were publicly not available so far—could further deepen the understanding about identifying the most promising regularized SPoC variants given a new data scenario.

Two important aspects can be transfered from the simulation to the real-world data scenario. First, the simulation allowed deriving an operating range of the regularization hyperparameter α for each CV-based regularization variant. When comparing these findings with the real-word data, we found that the optimal choice of the regularization intensity α for the CV-based techniques is in good accordance with the derived 'no-go' areas obtained from our simulations.

Second, according to the simulation under label noise in Fig. 4.6, we could gain an estimate of the label noise conditions ξ_n of any real-world dataset directly by comparing the absolute achievable decoding levels with the real-world decoding performances in Fig. 4.8. As an example, for the best performing subject of Fig. 4.8 (E) with z-AUC ≈ 0.78 on $N_{train} = 310$ data points, the label noise level can be estimated as $\xi_n \approx 0.2$ according to Fig. 4.6 (B). Despite such estimates may not perfectly represent the ground truth, they might be beneficial for comparing data from multiple experimental paradigms, e.g., in order to choose most suitable regularization strategies.

4.5.3 *CV-based vs. analytical model selection*

Overall, we introduced three CV-based Tikhonov regularization methods for SPoC (see overview in Tab. 4.1) and compared their performance against two variants based on automatic covariance shrinkage. Although the decoding performance of all three Tikhonov variants are on comparable levels, they strongly differ in terms of their sensitivity range for the regularization parameter. This information, however, is of great importance when it comes to choosing parameters by cross-validation. Interestingly, we found that NTik-SPoC and ASNTik-SPoC profit from a logarithmically scaled search space wrt. regularization parameter α while Tik-SPoC could also cope with a linear scaling. We conclude that this behavior is introduced by the additional trace normalization. When comparing NTik-SPoC and ASNTik-SPoC, the inclusion of additional LW-based shrinkage for the numerator regularization realized by ASNTik-SPoC does not boost performance significantly. Accordingly, NTik-SPoC seems preferable in a direct comparison due to its lower computational effort. In future work, an alternative data-driven estimation of the regularization parameter without cross-validation might be achieved, e.g., by utilizing a Bayesian framework which estimates the regularization strength via expectation maximization [194].

Comparing both LW-based covariance shrinkage based approaches, AS-SPoC seems to be the better choice compared to aTik-SPoC. Three arguments support this view. First, referring to the label noise challenged simulation in Fig. 4.6 we found that AS-SPoC profits from regularization under high label noise even for larger training sets ($N_{train} \gtrsim 300$) while this effect was less pronounced for aTik-SPoC. Second, we found that the analytically derived regularization parameter for aTik-SPoC across subjects is chosen way smaller compared to values chosen by CV for NTik-SPoC. For aTik-SPoC, the concatenation of epochs results in $N_s \cdot N_{train}$ sample points to estimate $\mathbf{\Sigma}_{avg}$. As the LW-based regularization parameter is anti-proportional to the number of samples [186, 187], an overly small regularization parameter is chosen, irrespectively of whether the covariance estimate did improve. Third, the analytic approach makes an i.i.d. assumption about the data. A violation thereof due to outliers might be compensated with a CV-based strategy but not by aTik-SPoC. The i.i.d. assumption might also be violated for AS-SPoC when the LW-based analytical solution for the trial-wise covariance matrix is challenged by autocorrelated data of a single epoch. A potential mitigation may be provided by alternative covariance shrinkage estimators that accounts for autocorrelated data as proposed by Bartz et al. [188]. Alternatively closed-form solutions for covariance shrinkage assuming elliptical distributions could also prove superior to the LW-based solution [195].

4.5.4 *Guidance for the practitioner*

Both, simulation and real-world data results strongly indicate that there is not one single regularization variant that outperforms all others. Different global parameters, such as dataset size, the noise conditions or non-stationarity in the data influence the achievable decoding accuracy.

The work by Engemann and Gramfort [196] reported the superiority of CV-based compared to analytical model selection in the context of spatial whitening of M/EEG data. This supports our proposal to prefer the CV-based approaches Tik-SPoC or NTik-SPoC over the LW-based AS-SPoC method. All three methods, however, are analytically solvable by an eigenvalue decomposition and require relatively low computational effort. As they may come up with partially disjunct components, we thus propose in practice to evaluate all three variants in parallel. The final feature set should be selected in a data-driven strategy to deduce the overall most relevant oscillatory components for a given application scenario.

4.6 LESSONS LEARNED

To sum up, this chapter provided novel regularization variants for SPoC and evaluated their characteristics in simulation and real-world data scenarios.

The simulation delivered two main results: First, it allowed comparing and explaining characteristics of the regularized SPoC algorithms. We could study the influence of varying training set sizes, label noise and

of the bandpower fluctuation width of the neural sources of interest. On the one hand, we found that the achievable overall decoding performance decays under increased label noise conditions and smaller datasets. On the other hand, small datasets and label noise were the settings under which several regularized SPoC variants could outperform the original unregularized algorithm. Second, the simulation outcomes offered a guideline for practitioners. It proposes to tune the search for a suitable regularization parameter to a log-scaled search space. Furthermore, it indicates that the number of training data points and label noise present in the data should guide the choice of this parameter.

As an additional validation, we tested the regularized SPoC algorithms on two real-world EEG datasets. Its outcome supported the guidelines obtained by simulation concerning the choice of regularization parameters and achievable performance improvements. We found, that varying datasets could profit strongly from different forms of regularization with increased decoding performances by up to 27 %. As a consequence, we recommend testing several versions of regularization if decoding performance is to be optimized in practice.

While we have chosen to compare relatively simple and general regularization techniques, this work could be expanded to more sophisticated regularization strategies, e.g., to realize session-to-session or subject-to-subject transfer scenarios. The presented regularization framework and the evaluation strategy using simulated and real-world datasets may pave this way.

5

MINING WITHIN-TRIAL OSCILLATORY BRAIN DYNAMICS OF SPATIALLY FILTERED SIGNALS

This chapter mainly builds upon the journal paper "Mining within-trial oscillatory brain dynamics to address the variability of optimized spatial filters" by Meinel, Kolkhorst, and Tangermann [34] © 2019 IEEE. In this collaborative work, I have taken the lead in developing the novel method, in validating and evaluating the evolved concept, in visualizing the results and writing the manuscript. For this thesis, the content has been extended by evaluating the proposed approach on data of a second experimental paradigm. Moreover, additional analysis aspects about the validation of the proposed mining framework are provided.

SUMMARY

Data-driven spatial filtering algorithms optimize scores such as the contrast between two conditions to extract oscillatory brain signal components. Most machine learning approaches for filter estimation, however, disregard within-trial temporal dynamics and are extremely sensitive to changes in training data and involved hyperparameters. This leads to highly variable solutions and impedes the selection of a suitable candidate for, e.g., neurotechnological applications. Fostering component introspection, we propose to embrace this variability by condensing the functional signatures of a large set of oscillatory components' into homogeneous clusters, each representing specific within-trial envelope dynamics. The proposed method is evaluated on two paradigms with a rich within-trial structure. For both scenarios, we found that the components' distinct temporal envelope dynamics are highly subject-specific and strictly confined regarding their underlying frequency band. As the analysis method is not limited to a specific spatial filtering algorithm, it could be utilized for a wide range of neurotechnological applications, e.g., to select and monitor functionally relevant features for BCI protocols in stroke rehabilitation.

5.1 INTRODUCTION

Spatial filtering algorithms have to deal with various challenges, such as the low signal-to-noise ratio of high-dimensional M/EEG recordings, non-stationarities over time and generally small training datasets [60, 90]. As a result, even a slight change of the training data can cause a large variability

of obtainable oscillatory features [31]. In addition, most approaches need to be configured by a set of algorithm-specific hyperparameters such as frequency bands, time intervals or regularization parameters, among others. Every hyperparameter set can result in different oscillatory features, such that applying a certain set may mean to miss relevant features. Conversely, modifying the hyperparameters actively bears the chance to detect other and even better features [197]. Overall, the selection of the (few) best hyperparameter set(s) in a large space calls for an optimization criterion.

So far, most BCI applications were mostly tuned to solely maximize the *decoding quality*. This criterion has been a good choice, if the final goal of the BCI system is to gain rapid and precise control. Using solely performance as an optimization criterion, however, does not consider the functional role of oscillatory features directly [90]. While a manual assessment of functional relevance is possible in small hyperparameter spaces, it turns to be impractical if they become large. Omitting any feature introspection, however, may be a missed opportunity as details of, e.g., ERD/ERS characteristic might provide an equally beneficial criterion from a clinical perspective [198].

Stepping beyond decoding accuracy, current literature reveals a limited amount of studies that specifically explore reliable and physiologically plausible EEG features (both within and across sessions and subjects) within the huge space of possible features. Typically, this can be achieved by means of clustering approaches [90], e.g., to extract homogeneous groups of spatial filters [156, 199], to identify EEG features encoding a similar stimulus response [200], to partition oscillatory features into groups of similar spatial, temporal and spectral properties [57, 59, 201–203] or to identify artifactual components [204]. Consequently, it would be desirable to consider these two aspects, decoding performance and the functional role of features, in a unified approach.

Research question Q2b

In this chapter, we contribute a novel data-driven approach to identify *reliable and functionally relevant* oscillatory features evaluated on two different motor task datasets. We hypothesize that the functional role of features can be assessed by the rich inner temporal envelope dynamics that is accompanied within a single trial along *multiple* events of a complex task. In an offline analysis performed on individual subjects, we first explore a large configuration space to embrace the variability of oscillatory features derived by a spatial filter approach. Then their event-related envelope dynamics are exploited by a clustering step to condense the large space to a small set of reliable oscillatory components that reveal homogeneous event-related envelope dynamics. Our approach finally allows a component selection that takes the functional role of oscillatory features into account. Hence, we provide a tool to practitioners that might enhance the efficacy of closed-loop interaction protocols, e.g., in the context of BCIs for stroke rehabilitation.

5.2 METHODS

5.2.1 *Datasets for evaluation*

For the evaluation of our method, the previously introduced datasets (D1a), (D1b) in Sec. 3.2.2 and (D2) in Sec. 4.3.1.1 were utilized. Both datasets comprise EEG recordings of a motor task but with different complexity regarding the underlying within-trial structure. Hereafter, we will shortly revisit the paradigms and explain their within-trial event structure which will subsequently be exploited by our mining approach.

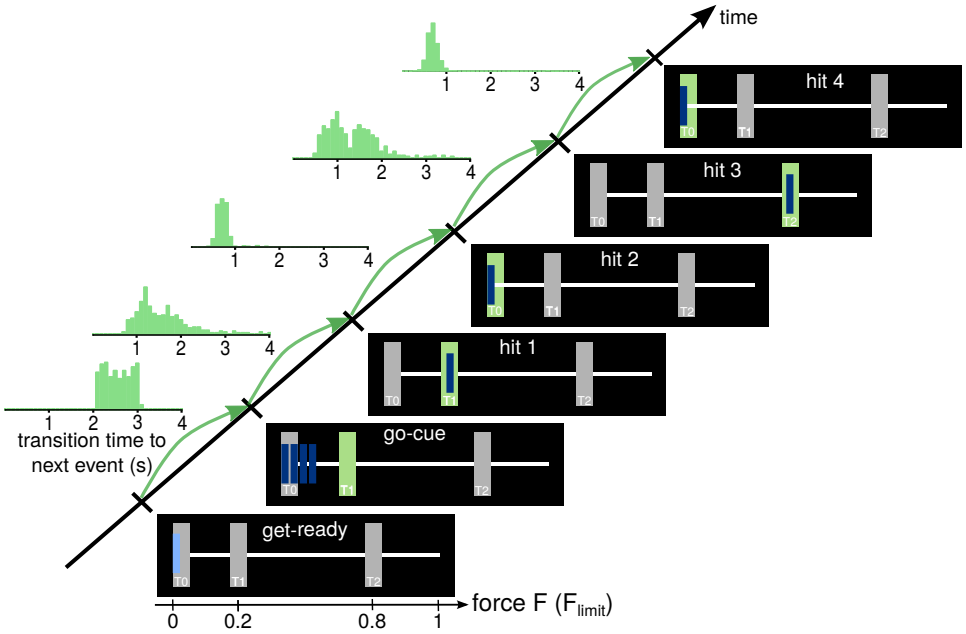


Figure 5.1: **Time course of a single SVIPT trial.** For each *within-trial* event, the user feedback on screen is shown (from bottom-left to top-right). Histograms left of the time axis indicate the distributions of transition times between events. The distances between events on the time axis are not realistically spaced. © 2019 IEEE

5.2.1.1 *SVIPT dataset*

As already described earlier in Sec. 3.2.1, SVIPT is a repetitive visuo-motor hand force task in which subjects are asked to control a horizontally moving cursor by a force transducer.

A single SVIPT trial traverses three stages and contains multiple within-trial events: the appearance of a light blue cursor initiates the *get-ready* phase with duration uniformly varied between 2 and 3 s. As sketched in Fig. 5.1, an inactive cursor appears on the leftmost border of the target field *T*0, while the force transduction into a horizontal cursor position is inactive. The subject was instructed to fixate the cursor and wait for a *go-cue* which was indicated by switching the cursor's color to dark blue. Starting with the *go-cue* event, the cursor can be horizontally displaced by applying force. Its rightmost position at F_{limit} is set to 30 % of the user's maximal force.

The *go-cue* marks the running phase of the trial in which the subject is asked to maneuver the cursor through a sequence of target fields as fast and as accurately as possible. A green shading visually indicates the current target field at any time during the *running* phase (see Fig. 5.1). Entering a target field with the cursor, a dwell time of 200 ms was required to evoke a successful hit event. The latter was indicated visually by a switch of the target field (another green-shaded target field appeared) or by the trial end. Each trial consisted of four hit events (*hit 1* to *hit 4*) and was randomly assigned to one of the two possible conditions, namely the specific target field sequence.

For the evaluation of our method, we utilized dataset (D1a) (for details see Sec. 3.2.2 and Tab. 3.1). Results on (D1b) are shown in the appendix Sec. A.3.

5.2.1.2 *Motor imagery dataset*

A typical use case for spatial filter optimization [53] is to classify between two or more classes in motor imagery (MI) datasets. Thus, the MI dataset (D2)—originally published by Blankertz et al. [190]—was utilized as previously introduced in Sec. 4.3.1.1. The dataset comprised single-session MI calibration recordings of the three classes *left hand*, *right hand* and *foot* with 75 trials per class. After preprocessing, data of 12 subjects remained with a minimum of $N_e = 100$ epochs.

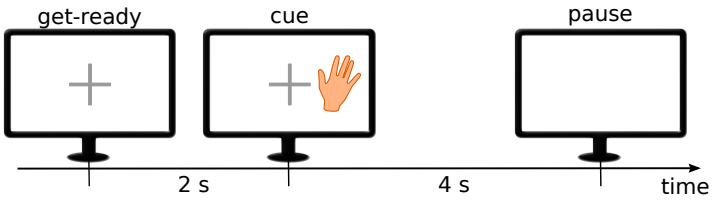


Figure 5.2: **Time course of a single MI trial.** For each *within-trial* event, the user feedback on screen is shown. The pause lasted 2 s in each trial.

As shown in Fig. 5.2, a single MI trial consisted of the three within-trial events: The appearance of a fixation cross on a screen indicated the *get-ready* event. Subjects were instructed to focus their view on the cross and wait for an upcoming cue. After 2 s, the *cue* event was visually displayed by an arrow pointing either to the left, to the right or downwards. The subject was instructed to perform the corresponding motor imagery task for 4 s before the arrow vanished on screen and thereby indicated the *pause* event. After 2 s of rest, the next trial started. Overall, the transition time between events was constant throughout the full session which was not the case for SVIPT as presented in Fig. 5.1.

5.2.1.3 *EEG data preprocessing*

The single-session EEG data of the datasets (D1a,b) and (D2) underwent the same offline preprocessing. It comprised of low-pass filtering the raw data at 100 Hz, sub-sampling to 500 Hz sample rate before high-pass filtering

at 1 Hz. For frequency filtering, linear Butterworth filters of 5^{th} order were applied. Noisy channels were removed by a two-step procedure. First, the variance of single epochs and channels was computed. Based on the pooled statistics, all cases outside the $[10, 90]$ percentiles and also exceeding twice the corresponding inter-percentile range were registered as outliers. Second, channels which allocated more than 10% of all outliers and a minimum of 5% outliers across epochs were removed. Furthermore, artifact cleaning was done by an independent component analysis (ICA) decomposition on data of the first run of each session. The ICs were rated for artifactual origin with the automated artifact detection framework MARA [89]. Based on MARA's probability ratings, a conservative criterion was applied by removing only up to 10 most likely artifactual ICs from the data before projecting the data back into the original sensor space. Only this pre-cleaned data was used in the next steps.

5.2.2 *Optimized spatial filtering for single-trial EEG analysis*

A variety of optimization criteria for spatial filter estimation can be found in literature (see Sec. 2.3.4). Typical algorithmic solutions for unsupervised scenarios include PCA or ICA. In the context of BCIs, supervised algorithms such as CSP [93] or SPoC [94] are state-of-the art approaches, for which also a large number of variants have been proposed. Regardless of the specific algorithm, a set of hyperparameters needs to be determined, which influences the spatial filter optimization. Prominent examples are the choice of the training dataset, of a subject-specific frequency band or to select components up to a particular rank. In general, choosing different sets from the large search space of possible hyperparameter configurations typically leads to different estimated filters.

In the following, we present a method to identify consistent spatial filters within a non-trivial hyperparameter space. Even though this approach can be *utilized for any spatial filtering method*, we exemplarily show and evaluate its application for two different spatial filtering variants each applied on data of a different paradigm. The resulting two decoding scenarios are shortly introduced hereafter.

5.2.2.1 *Hyperparameter space for oscillatory component analysis*

Decoding scenario (1): NTik-SPoC analysis on SVIPT data

As a first application scenario, we chose the dataset (D1a) and (D1b) to decode trial-wise reaction time from multi-channel EEG recordings segmented relative to a time interval close to the SVIPT *go-cue*. This decoding scenario has previously been described in Secs. 3.2.5 and 4.3.2.1. Following the results of Chap. 4, we decided for the overall favorable NTik-SPoC algorithm of Sec. 4.2.1 which includes a Tikhonov regularization and trace normalization of two involved covariance matrices. For the computation of a single spatial filter $\mathbf{w}$, a set of different hyperparameters is involved.

To explore the high variability in trained decoding models, the following hyperparameters were considered for the SVIPT decoding scenario:

1. In the temporal domain, initial data segmentation into N_e epochs requires the definition of a time interval $[t_0, t_0 + \Delta t]$ with starting time point t_0 and interval length Δt. For motor performance decoding in SVIPT, the recorded EEG data were segmented relative to the *go-cue* event with fixed $\Delta t = 1\,s$. Time point t_0 was chosen among the values $\{-1, -0.75, -0.5\}\,s$.

2. The frequency domain is characterized by the central frequency f_0 and bandwidth Δf. Overall, 45 exponentially increasing and overlapping frequency bands with $f_0 \in [1, 95]$ Hz and $\Delta f \in [2, 10]$ Hz were evaluated.

3. The utilized spatial filtering method generally provides a full-rank decomposition, thus the rank k can be seen as another hyperparameter for thresholding the composition. We considered spatial filters of the first $k = 1, ..., 8$ ranks for further analysis.

4. NTik-SPoC requires the tuning of the regularization strength α. Based on findings reported in [32, 33], choosing the regularization strength $\alpha \in [10^{-8}, 10^{-3}]$ allows outperforming non-regularized SPoC wrt. decoding accuracy (see Sec. 4.4).

5. Upon each hyperparameter configuration, a 5-fold chronological cross-validation procedure was employed for the calculation of a spatial filter. As many spatial filter optimization problems can be formulated as a generalized eigenvalue problem, a set of filters $\{\mathbf{w}_{k,q}\}$ can be derived from each fold q, with k corresponding to the rank in the decomposition.

In summary, the different hyperparameter configurations span the configuration space $\Omega = \{(t_0, f_0, \Delta f, k, \alpha, q)\}$ with an overall number of $|\Omega| = 81,000$ configurations. We will refer to a single configuration by $\omega_j \in \Omega$ for $j = 1, ..., |\Omega|$. It determines a single spatial filter calculation. Hereafter, every spatial filter $\mathbf{w}$ corresponds to a single configuration ω_j.

Decoding scenario (2): CSP analysis on MI data

As a second decoding scenario, a regularized version of the CSP algorithm using automatic shrinkage [62] is deployed to compute spatial filters which allow distinguishing between *left hand* and *right hand* trials in the motor imagery dataset (D2). Based on a $1\,s$ time interval after the *cue* event, CSP was trained with the following hyperparameters involved:

1. In the temporal domain, data was segmented relative to the *cue* event with fixed interval length $\Delta t = 1\,s$. Time point t_0 was chosen among 20 values in the interval $[0, 5]\,s$.

2. Regarding the frequency domain, the same configuration space as before with logarithmically increasing and overlapping frequency bands ranging from $\approx 1 - 100\,\mathrm{Hz}$ were evaluated.

3. As CSP also provides a full-rank decomposition, we selected the first four ranks k from both ends of the eigenvalue spectrum.

4. In analogy to scenario (1), a 5-fold chronological cross-validation procedure was employed for filter estimation. Thus, a set of filters $\{\mathbf{w}_{k,q}\}$ can be derived from each fold q.

In summary, the configuration space Ω comprised $|\Omega| = 36,000$ configurations. Compared to scenario (1), the search space is barely halved but the fraction of resulting informative configurations is almost doubled in scenario (2) (see results in Tab. 5.1). Thus, the resulting configuration spaces for both scenarios after denoising (see Sec. 5.2.3.2) were comparable.

5.2.2.2 *Evaluation scheme for spatial filter computation*

Both decoding scenarios were evaluated within a 5-fold chronological cross-validation procedure such that the data is split in train $\{\mathbf{X}_{tr}, z_{tr}\}$ and test sets $\{\mathbf{X}_{te}, z_{te}\}$. After estimating the labels z_{est} on different test data points $\mathbf{x}_{te}$, the decoding performance of a single filter $\mathbf{w}_{tr}$ was assessed by the z-AUC metric (see Sec. 3.2.7) for the regression models trained in scenario (1) and by the AUC metric [141] for the classification models in scenario (2). Both measures characterize the separability of the estimated labels when comparing them to the known test labels z_{te}. A perfect decoding is reflected by a value of 1, while chance level is reflected by 0.5. Hereafter, we will omit the train/test subscripts to simplify the notation.

5.2.3 *Method for mining oscillatory components*

Our method to identify groups of functionally relevant oscillatory components is depicted in Fig. 5.3. It consists of three steps: first, oscillatory components from EEG data of a single subject are computed across a large hyperparameter space. This results in a broad variety of spatial filter examples and embraces the variability of the decoding method. Details on the screened configuration space are given in Sec. 5.2.2.1. Second, the large component space is de-noised by restricting the analysis to *reliable* components via selection criteria on single components. Third, for all remaining components the event-related envelope dynamics of each source component is exploited by a clustering step. Therefore, envelope features are preprocessed and condensed. The goal is to finally identify clusters of oscillatory components which reveal distinct and stable envelope dynamics.

In the following, details on the three steps are given. The Python code of the method is accessible on GitHub[1] and is partially based on the machine learning library *scikit-learn* [205].

1 https://github.com/bsdlab/func_mining

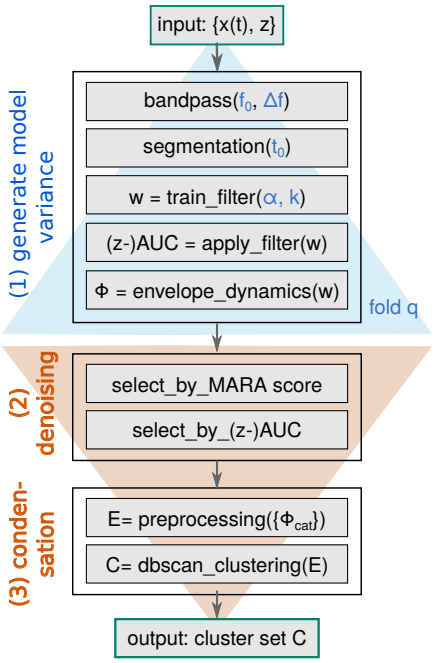


Figure 5.3: **Pseudo-code scheme for mining envelope dynamics.** Subject-specific data $x(t)$ and labels z enter part (1) to sample oscillatory components in a large hyperparameter space and thus sample variable spatial filter solutions. In blue, all included hyperparameters for spatial filter $\mathbf{w}$ estimation are highlighted. Part (2) is a denoising step by reducing the overall configuration space before step (3) condenses envelope features into clusters.

5.2.3.1 *Extracting envelope dynamics*

The temporal dynamics of an oscillatory source component $\mathbf{w}$ based on configuration ω_j, given sensor data $\mathbf{x}(t)$, can simply be derived via the backward model $\hat{s}(t) = \mathbf{w}^\top \mathbf{x}(t)$. It requires to bandpass filter the data $\mathbf{x}(t)$ to the same frequency band $[f_0 - \frac{\Delta f}{2}, f_0 + \frac{\Delta f}{2}]$ on which the spatial filter $\mathbf{w}$ has been trained initially. Hereto, zero-phase linear Butterworth filters of 5^{th} order were applied. As an alternative to the variance approximation, the envelope time course $\phi_j(t)$ can be estimated by the magnitude of the analytic signal which is given by the Hilbert transformation $\mathscr{H}(\cdot)$ on narrow-band data:

$$\phi_j(t) = |\mathscr{H}(\hat{s}(t))| = |\mathscr{H}(\mathbf{w}^\top \mathbf{x}(t))| \tag{5.1}$$

The component-specific envelope dynamics provide a rich source of information by integrating spatial, spectral and temporal aspects of the multivariate data of individual subjects. Precisely, we analyzed within-trial event-related envelope dynamics and thus exploit the inner event structure of an underlying experimental paradigm. Hereafter, the paradigm-specific feature extraction steps are described for both selected scenarios.

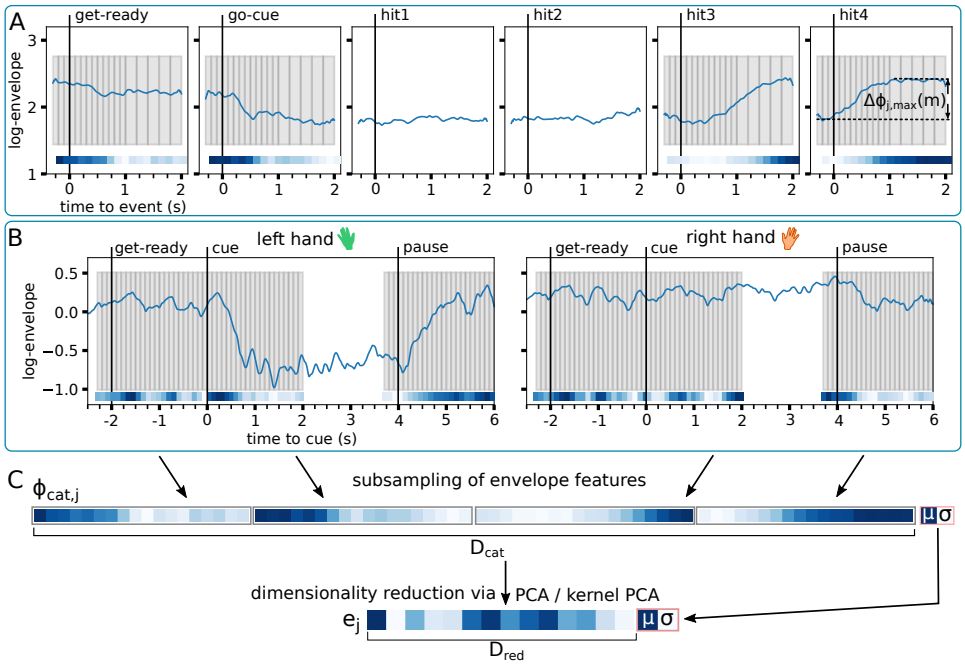


Figure 5.4: **Feature extraction scheme based on envelope dynamics of a single component.** The session average log-envelope of an oscillatory component was epoched and aligned to the various within-trial events occuring in SVIPT (A) or the MI paradigm (B). In gray, time intervals for temporal subsampling of the event-related time series are depicted. (C) After concatenating the subsamples of the most relevant events (details see text), the resulting feature vector $\phi_{cat,j}$ was standardized to zero mean and unit variance as illustrated by the color coding. After dimensionality reduction, 10 condensed features plus the two standardization features form the final feature vector for clustering.

Decoding scenario (1): NTik-SPoC analysis on SVIPT data

For the SVIPT dataset (D1a), each spatial filter $\mathbf{w}$ was solely trained on data segments extracted from the vicinity of the *go-cue* event. However, as depicted in Fig. 5.1, a single SVIPT trial also contains a rich inner structure of multiple events m (*get-ready*, *go-cue* and *hit 1* to *hit 4*). We expect that this within-trial structure is at least partly captured by the temporal envelope dynamics of single oscillatory components. In other words, we test for the generalization strength of $\mathbf{w}$ on unseen data. Therefore, the envelope time course of each configuration j was segmented in a time interval $t \in [-300, 2000]$ ms relative to each event m and averaged across the $N_e = 400$ epochs. This resulted in $\phi_j(t, m)$ as exemplarily sketched for one component in Fig. 5.4 (A).

Decoding scenario (2): CSP analysis on MI data

For the MI dataset (D2) and the corresponding CSP analysis, all components were trained exclusively upon a 1 s time interval between the *cue* and *pause* event (see blue shaded time intervals in Fig. 5.6). To test the generalization strength of the optimized CSP filters, we segmented the component's

log-envelope along the included events (*get-ready*, *cue* and *pause*) and in addition split the data by the corresponding class label (see Fig. 5.4 (B)). The segmentation interval was identical to the one of scenario (1), while a maximum of $N_e = 75$ epochs was available per class.

5.2.3.2 *Denoising clustering input data*

The overall spanned hyperparameter space Ω elicits a large variation of derived spatial filters $\mathbf{w}$ of which a large fraction might be non-stable or of artifactual origin. To control for their reliability, the overall component space was reduced by deploying two hard selection criteria for both decoding scenarios: first, each spatial filter is expected to exceed a robust decoding performance level above chance. Therefore, each spatial filter is required to result in minimum z-AUC/AUC decoding performance above a threshold level of 0.6 on test data. The threshold was determined on a group-level analysis [28]. Second, the neural origin of a spatial filter is verified by applying MARA [89], an automatic classification tool to distinguish between neural and artifactual components, including ocular or muscular activity. It returns a posterior artifact probability p_{art} for a given single oscillatory component. To restrict the resulting component space to mostly neural components with a high certainty, we required a probability of $p_{art} \leq 10^{-5}$ for each component.

In summary, the two selection steps reduced the original configuration space Ω to a subject-specific subset $\Omega_{sel} \subset \Omega$ which comprises all hyperparameter configurations that survived the component selection. Hereafter, the configuration index j refers to the set Ω_{sel}.

5.2.3.3 *Features for clustering*

For each configuration $\omega_j \in \Omega_{sel}$, the corresponding event-related envelope dynamics $\phi_j(t)$ were extracted for both scenarios as depicted in Fig. 5.4 (A) and (B). As clustering methods generally depend on the evaluation of a distance metric which becomes unreliable for high-dimensional spaces, a widely established strategy is then to reduce the dimensionality of the input feature space [206] prior to clustering. Therefore, two steps were taken:

1. *Subsampling and standardization:* the session-averaged envelope time series aligned to multiple events—and also classes in scenario (2)—were log-scaled and temporally subsampled as schematically shown in Fig. 5.4 (A) and (B). The log-scaling was applied to obtain approximately normally distributed features. This step resulted in a concatenated feature vector $\phi_{cat,j} = \phi_{cat}(\omega_j) \in \mathbb{R}^{D_{cat}}$ with $D_{cat} = 72$ for scenario (1) and $D_{cat} = 66$ for scenario (2). For the sake of comparable feature dimensionality across scenarios, a larger subsampling interval for the SVIPT scenario was chosen for the late phase of events as less variations in time were observed here. After computing mean μ and standard deviation σ of a feature vector $\phi_{cat,j}$, it was standardized to zero mean and unit variance. Overall, applying this procedure

to all selected configurations $\omega_j \in \Omega_{sel}$, this yielded a subject-specific dataset $\Lambda = \{\phi_{cat,j}\}$.

2. *Dimensionality reduction:* on the concatenated feature set Λ, a dimensionality reduction step was performed by either PCA or kernel PCA to $D_{red} = 10$ components. The performance of both methods will be compared in the results (see Fig. 5.9). Kernel PCA was computed by a radial basis function kernel with $\gamma = 1/D_{red}$. The final feature vector e_j was composed of the 10 subspace features based on Λ and two additional features of the standardization step (μ and σ) resulting in $e_j \in \mathbb{R}^D$ with $D = 12$ dimensions. Overall, a subject-specific dataset $E = \{e_j\}$ of $|\Omega_{sel}|$ samples was obtained.

5.2.3.4 *Clustering algorithms*

To find sets of components with homogeneous envelope dynamics, we aim to group them by searching non-overlapping clusters of components. Assuming that a rich within-trial envelope structure is only expected for a small fraction of all configurations, it may not be necessary to assign each configuration ω_j to a cluster. For more details on the general concept of clustering, we refer to the review by Jain [207]. Let a clustering of E return a set of disjoint clusters that splits E into $|C|$ groups with $C = \{c_k\}$ and $k = 1, ..., |C|$.

Density-based clustering

For partitioning the dataset E, the DBSCAN [208, 209] algorithm was utilized, which realizes a density-based clustering. DBSCAN groups dense regions of a dataset to clusters by checking for every sample if: (1) at least m_{pts} other samples are in ϵ range to this sample or (2) at least one neighboring sample in ϵ distance is enclosed. In the first case, the sample is called a *core sample* while in the second case it is referred to as a *border point*. If none of the two criteria are fulfilled, the sample receives an *outlier* label. As such, the density-based definition of a cluster requires that each cluster sample reaches *at least one* other sample in ϵ distance. DBSCAN does not make any assumption on the cluster shape, thus it provides the possibility to identify non-linearly separable clusters. In this thesis, DBSCAN was evaluated on Euclidean distances $d_{euc}(e_i, e_j)$ between samples e_i and e_j in a condensed envelope feature space.

DBSCAN involves two hyperparameters, m_{pts} and ϵ. Regarding m_{pts}, we followed the suggestion of Sander et al. [210] and took the feature dimensionality D into account by setting $m_{pts} := 2D$.

The choice of ϵ is the more sensitive parameter, as the number of clusters $|C|$ diminishes strongly for an increased ϵ [209]. Overall, we expected that the envelope dynamics would not reveal a rich structure for all hyperparameter configurations $\omega_j \in \Omega_{sel}$, e.g., if ERD/ERS effects are not present. Hence, we expected rather large outlier clusters. Typically, in many clustering scenarios it is sufficient to simply maximize the average silhouette

score S [211] which captures the ratio of within-cluster homogeneity to the closest neighboring cluster (definition see Eq. (5.4) below). But then the average silhouette S across all samples would most probably be dominated by the outlier class. Thus instead of using the silhouette score to guide the clustering, we aimed to obtain the maximal number of homogeneous clusters N_{hom}:

$$N_{hom}(\epsilon) = \sum_{c_k \in C(\epsilon)} \Theta(\min_{e_i \in c_k}(S(e_i, c_k)) \geq S_{hom}) \tag{5.2}$$

with the unit-step function $\Theta(x) = 1$ for $x \geq 0$ and $\Theta(x) = 0$ for $x < 0$. N_{hom} refers to the total number of clusters and the sample e_i with smallest silhouette $S(e_i, c_k)$—as defined in Eq. (5.4)—of each cluster c_k is required to exceed a threshold silhouette score S_{hom}. This translates to the following optimization criterion for ϵ:

$$\epsilon^* = \underset{\epsilon \in [\epsilon_{min}, \epsilon_{max}]}{\arg\max}\ N_{hom}(\epsilon) \tag{5.3}$$

For this thesis, a subject-independent threshold of $S_{hom} = 0.2$ was set in order to allow for comparable clustering results such as $|C|$ per subject. This threshold value enforces a slightly smaller within-cluster distance compared to the nearest neighbor distance. The interval $[\epsilon_{min}, \epsilon_{max}]$ was automatically determined by an ordered k-distance plot which reports the k-th nearest neighbor distance (NND) for each sample in E. The distances are arranged in ascending order starting with the smallest distance. This procedure was described by [208, 209] together with setting $k = 2D - 1$. For ϵ_{max}, we detected the first substantial increase of the k-distance plot (starting with smallest distances) by a variance criterion in order to find the end of the "valley" of lowest distances. As lower boundary, ϵ_{min} was determined by the 2^{nd} percentile of the D-th NND distribution. Finally, ϵ was evaluated at 60 values from the subject-specific interval $[\epsilon_{min}, \epsilon_{max}]$. In principle, the DBSCAN approach is of deterministic nature. With a permutation of a dataset this characteristic may change in some cases as then samples might be assigned to different clusters induced by the switched order of the dataset [209].

K-means clustering

As a baseline method, the k-means algorithm [212] was also applied to validate the approach of clustering subject-specific envelope data E into $|C|$ clusters. In contrast to DBSCAN, the number of clusters $|C|$ is a hyper-parameter of the k-means approach and thus is required to be specified a-priori. The hyperparameter selection was accomplished by the same optimization criterion as for the DBSCAN algorithm (see Eq. (5.3)). However, $|C|$ replaced ϵ and was evaluated in an interval $|C| \in [2, ..., 2D]$. k-means partitions a dataset such that minimal intra-cluster variance is obtained. To overcome the algorithm's sensitivity to the initialization of the cluster centroids, an extension of the k-means seeding strategy was used — known as *k-means ++* [213]) — by initializing along far apart centroids.

5.2.3.5 *Validation metrics for clustering*

Given the condensed log-envelope dataset of different oscillatory components, the ground truth cluster labels are unknown. Thus we require a label-free validation metric [214, 215] to judge the quality of the clustering outcome. Three different categories of validation metrics are considered [211] here: (1) internal validation metrics based on features used for the clustering, (2) external validation metrics which utilize features that have not been used for the partitioning step and (3) context-specific metrics which are defined by domain-knowledge given specific characteristics of the oscillatory component datasets. In the following, for each metric a reference symbol is given in brackets as well as an arrow that indicates the direction towards a more preferable clustering (e.g., ↑ denotes "higher is better"):

- Silhouette score $S(c_k) \uparrow$ [216]: Given an arbitrary sample e_i assigned to cluster c_k, this internal clustering validation score relates the within-cluster similarity $a(e_i, c_k) = |c_k|^{-1} \sum_{e_j \in c_k \setminus \{e_i\}} d_{euc}(e_i, e_j)$ to the nearest neighbor dissimilarity $b(e_i, c_k)$:

  $$b(e_i, c_k) = \min_{c_l \in C \setminus c_k} [|c_l|^{-1} \sum_{e_j \in c_l} d_{euc}(e_i, e_j)]$$

  The silhouette score $S(e_i, c_k)$ for a sample e_i belonging to cluster c_k is defined as:

  $$S(e_i, c_k) = \frac{b(e_i, c_k) - a(e_i, c_k)}{\max[a(e_i, c_k), b(e_i, c_k)]} \tag{5.4}$$

  It can be verified easily that $-1 \leq S(e_i, c_k) \leq +1$, such that a "perfect" clustering will return a value of +1. Among other metrics, the silhouette metric has been shown to serve as a reliable *internal* clustering validation method—computed upon features used for the clustering—for various classes of clustering algorithms [211, 217]. Furthermore, the within-cluster silhouette score $S(c_k) = |c_k|^{-1} \sum_{e_i \in c_k} S(e_i, c_k)$ can be determined with low computational effort.

- Intra-cluster mean squared error IC-MSE ↓: For the envelope clustering, a number of preprocessing steps were applied in order to reduce the dimensionality of the original time resolved event-related envelopes (see Fig. 5.4). To verify the intra-cluster homogeneity of cluster c_k upon the original event-wise log-envelope time series, the event-specific mean $\phi_{avg}(c_k, t, m) = |c_k|^{-1} \sum_{\omega_j \in c_k} \phi_j(t, m)$ for cluster c_k was utilized to compute the mean squared error across the full set of events M and time samples T:

  $$\text{IC-MSE}(c_k) = \beta^{-1} \sum_{\omega_j \in c_k} \sum_{m \in M} \sum_{t \in T} (\phi_j(t, m) - \phi_{avg})^2 \tag{5.5}$$

  with $\beta = |c_k||M||T|$. To summarize, IC-MSE corresponds to the intra-cluster envelope variance. It can be seen as an external validation

metric as it is based on envelope features which were unseen by the clustering.

- Intra-cluster central frequency variation std $\left(f_0(c_k)\right)$ $\downarrow$: A context-specific validation is accessible via the central frequency hyperparameter f_0 which is involved in the spatial filter optimization. When capturing the within-cluster variation of f_0 by std $\left(f_0(c_k)\right)$, this value was expected to be rather small as EEG features are confined with respect to their spectral occurrence.

- Intra-cluster pattern heterogeneity ICPH: As a context-specific characterization metric, the spatial activity pattern $\mathbf{a}_j$ for each sample e_j of a cluster c_k was computed. As a measure of the within-cluster heterogeneity of spatial activity patterns for a cluster c_k, the cosine angle θ as defined in [32] between each $\mathbf{a}_j$ to a cluster representative pattern $\mathbf{a}^*(c_k)$ was averaged as follows:

$$\mathrm{ICPH}(c_k) = |c_k|^{-1} \sum_{e_j \in c_k} \theta(\mathbf{a}_j, \mathbf{a}^*(c_k)) \tag{5.6}$$

  The representative pattern $\mathbf{a}^*$ is defined by identifying the sample $e^* \in c_k$ with minimal Euclidean distance wrt. clustering features to all other samples of the same cluster.

- Event-specific maximal envelope difference $\Delta\phi_{max}(m, c_k)$: This metric serves to functionally characterize single clusters by their underlying ERD/ERS dynamics. As exemplarily sketched for *hit 4* in Fig. 5.4, $\Delta\phi_{j,max}(m)$ represents the maximum envelope difference across time within an event m and for a single configuration $\omega_j \in c_k$ of a cluster c_k. The value is referenced to the average log-envelope in the interval 500 ms prior to the event m. Averaging across all cluster configurations reveals the event-specific maximal within-cluster logarithmic envelope differences $\Delta\phi_{max,avg}$:

$$\Delta\phi_{max,avg}(m, c_k) = |c_k|^{-1} \sum_{\omega_j \in c_k} \Delta\phi_{j,max}(m) \tag{5.7}$$

  Given a homogeneous cluster, $\Delta\phi_{max,avg}(m) < 0$ refers to an ERD effect for event m, while $\Delta\phi_{max,avg}(m) > 0$ describes an ERS effect.

5.2.3.6 *Evaluation scheme for clustering step*

In this thesis, we evaluated our proposed methodology for two above described decoding scenarios (1) and (2). Therefore, the SVIPT datasets (D1a) and (D1b) as well as the MI dataset (D2) were utilized. For each subject of the corresponding dataset, the size of the selected configuration space $|\Omega_{sel}|$ was substantially different (see results in Sec. 5.3.1.4). To ensure comparability of clustering runs across subjects, we randomly sampled $N = 2000$ feature vectors e_j from each subject-specific dataset E while keeping the original sample order—to ensure DBSCAN to be deterministic—before entering the clustering. This procedure was repeated 12 times per subject.

5.3 RESULTS

First, the most relevant findings for practitioners (e.g., in the field of BCIs for rehabilitation) are presented. Here, representative subject-specific clusters of oscillatory components are described, and a way to *functionally* characterize the grouped event-related envelope dynamics. Thereby, the applicability of the method is demonstrated. Second, our approach is validated and characterized by presenting a group level analysis for the two different decoding scenarios. Unless noted differently, all results shown were achieved upon kernel PCA preprocessing.

5.3.1 *Findings for the practitioner*

For the practitioner, it may be of particular interest to gain understanding, which insights our method can provide on the level of single subjects and on the group level.

5.3.1.1 *Envelope dynamics of clusters*

Decoding scenario (1): NTik-SPoC analysis on SVIPT data

After applying the described within-subject approach on NTik-SPoC components computed for dataset (D1a), the original (non-condensed) within-trial event-related envelope dynamics of subject-specific clusters in Fig. 5.5 are reported. For all examples, the spatial filter and corresponding activity pattern of the cluster representatives—obtained by selecting the sample with minimal Euclidean distance in the feature space—are shown in columns (G) and (H). Rows (C1)–(C7) present different instances of exemplary clusters. They were chosen to represent a broad range of typically observed effects in terms of band-specific amplitude modulations, underlying frequency ranges and cluster homogeneity. Specifically, rows (C1)–(C3) refer to clusters of subject S13, while (C4)–(C5) are gained from subject S7 and (C6)–(C7) correspond to S5.

Considering the transition times between single events as provided by Fig. 5.1, the time between *hit 1* and *hit 2* as well as between *hit 3* and *hit 4* on average was around 800 ms. Thus, the event-locked envelopes in (C) and (D), respectively (E) and (F), contain overlapping information.
Based on Fig. 5.5, the following observations can be reported:

First, regarding all shown examplary clusters, the envelope dynamics aligned to the different within-trial SVIPT events reveal distinct and time-locked ERD or ERS effects which can be separated well by the clustering approach. For the cases reported in Fig. 5.5, ERD effects dominate for *get-ready* and *go-cue* events, while *hit 3* and *hit 4* elicit ERS effects.

Second, the displayed examples demonstrate that the event-related envelope dynamics reveal substantially different shapes both within and across subjects. Taking a closer look at the selected cluster instances of S13 with clusters (C1)-(C3) at the *get-ready* event, there are different effects visible. While components grouped into (C2) and (C3) reveal a slight step-like

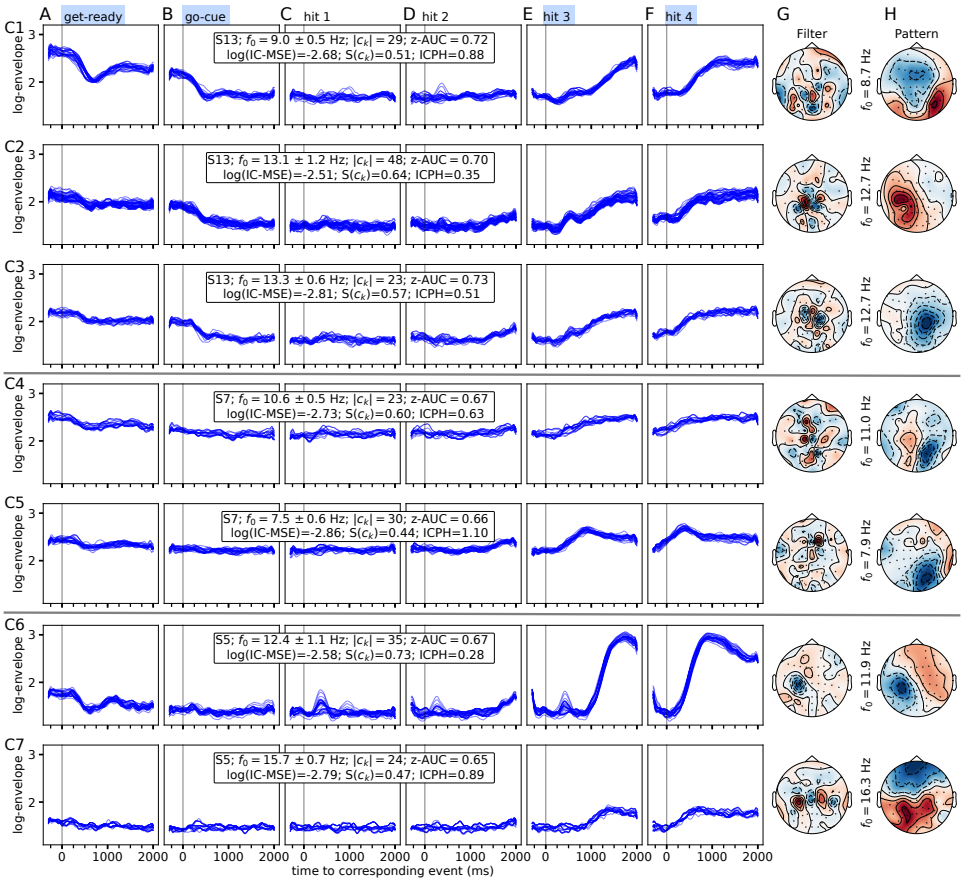


Figure 5.5: **Representative event-related envelope dynamics of single subject-specific clusters.** In rows (C1)–(C7), the corresponding envelope dynamics of all hyperparameter configurations for single subject-specific clusters c_k are reported. Columns (A)–(F) report the cluster-wise envelope dynamics for within-trial SVIPT events, while in (G) the spatial filter and in (H) the related activity pattern of cluster representatives (with annotated central frequency) are shown. In all subplots of columns (A)–(F), every blue line refers to the log-envelope dynamics $\phi_j(t, m)$ of one single hyperparameter configuration $\omega_j \in c_k$. Only events highlighted in blue were included for the clustering step. The text box on top of each row provides the subject code, the mean and standard deviation of the central frequency across all cluster samples, the cluster size, the average decoding performance as well as three validation metrics.

behavior, cluster (C1) comes with a strong ERD followed by an ERS effect. Regarding *hit 4* of (C6) and (C7), the examples for subject S5 nicely illustrate that amplitude differences can be substantially different across clusters and configurations. While clusters (C4) and (C5) of subject S7 are characterized by an ERS effect time-locked to *hit 3*, all remaining examples reveal the ERS at *hit 4*.

Third, we can report under which conditions our clustering approach works best, e.g., when comparing the cases (C2) and (C3). They nicely demonstrate that smaller cluster sizes correspond to more homogeneous clusters as visually observable and documented by, e.g., the IC-MSE values.

Fourth, similar silhouette and IC-MSE scores do not directly imply a high pattern homogeneity (low ICPH values), as can be seen comparing clusters (C1) and (C6).

For completeness, exemplary clusters obtained on dataset (D1b) of various chronic stroke patients are reported in the appendix A.3. Interestingly, also for this patient dataset with remarkably less trials per session, the proposed approach allowed to identify patent-specific clusters of homogeneous but still distinct envelope dynamics.

Decoding scenario (2): CSP analysis on MI data

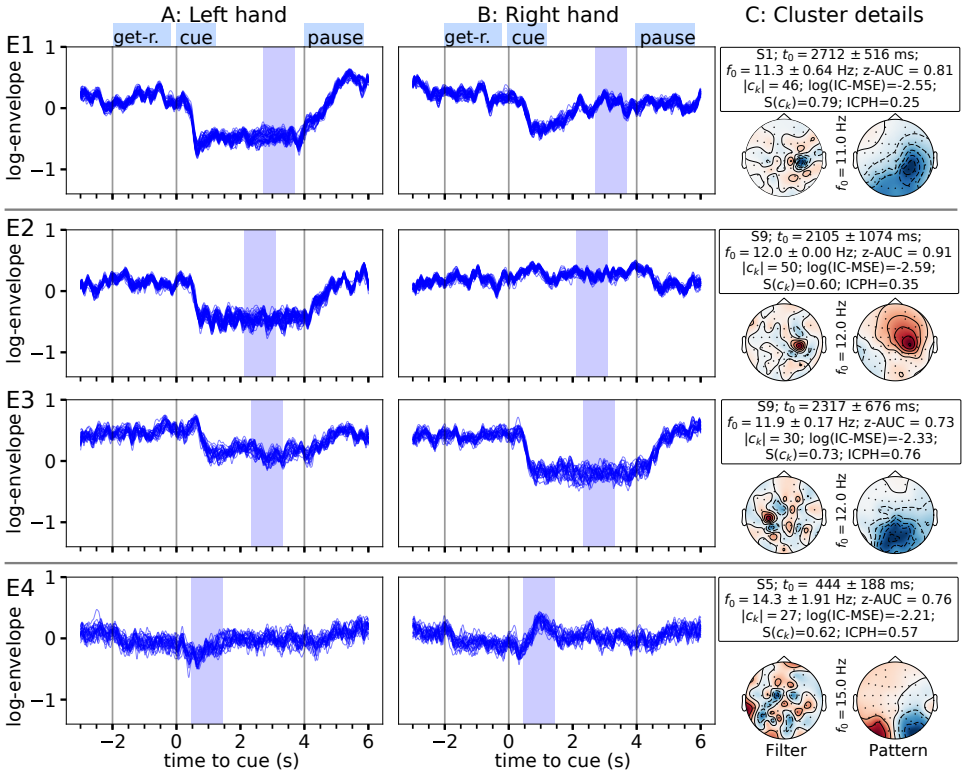


Figure 5.6: **Event-related envelope dynamics for motor imagery data.** The envelope dynamics of exemplary CSP model clusters are shown in rows (C1-C4) separated by *left* (A) or *right* (B) hand labeled epochs. Time $t = 0$ refers to movement onset. In column (C), the subject code, the average time and frequency parameters of the CSP models, the cluster size and thee validation metrics of the corresponding cluster are given. In addition, the spatial filter and pattern of the cluster representative are shown. For each cluster, the cluster-average time interval for the training of the CSP model is highlighted in blue.

In analogy to the structure of Fig. 5.5, rows (E1)–(E4) of Fig. 5.6 present representative cluster examples based on CSP components of the MI dataset (D2). The columns (A) and (B) reveal the class-wise envelope dynamics, while for each class three events were considered for the clustering step. The results on the MI dataset in Fig. 5.6 nicely indicate that our method allows for an individual assessment of envelope dynamics. This is demonstrated by

cluster examples (E1) and (E2), which originate from two different subjects. Both representative components live in a comparable frequency domain, and their patterns are highly similar reflecting a well-known lateralized motor component [53]. Interestingly, the underlying envelope dynamics of the two clusters differ substantially. While for (E2) there is an ERD only for *left hand* trials, there is an ERD for both classes observable in example (E1). Moreover, (E1)-(E3) capture alpha-band associated clusters, (E4) reports on a beta-band related cluster which shows already an ERD effect for the *get-ready* event. Observable in all given examples, the envelope dynamics among the two classes differs substantially, which is only partially enforced by the CSP algorithm, namely along the 1 *s* training interval.

5.3.1.2 *Functional assessment of clusters*

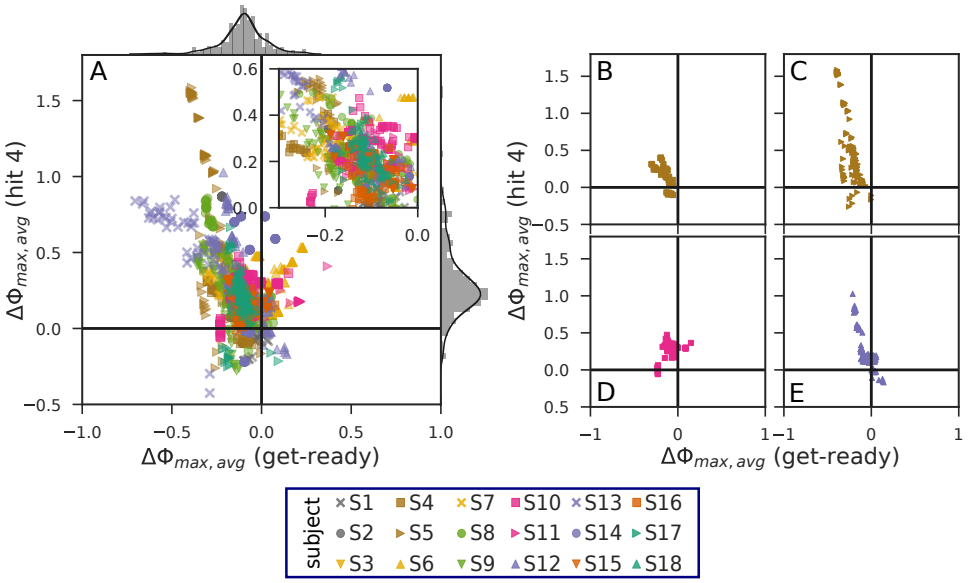


Figure 5.7: **Individual cluster-specific ERD/ERS effects exemplarily for two specific SVIPT events.** (A) The plot is based on pooled data from all clusterings across subjects. A single data point refers to an individual cluster and is encoded by its underlying subject. The inset plot shows a magnified version of the central area. Attached to both axes, the corresponding distributions across all displayed configurations are shown. Plots (B)–(E) show exemplary scatter plots for single subjects. © 2019 IEEE

Cluster-specific ERD/ERS intensities along the within-trial events indicate the functional role of the contained components and thus provide a way to characterize single clusters post-hoc. An example is given in Fig. 5.7 (A). It exemplarily describes the logarithmic envelope differences within *get-ready* and *hit 4* events.

The distributions on the x- and y-axis reveal ERD effects time-locked to *get-ready* and *go-cue* (not shown here) and a subsequent ERS effect with the last hit event per trial. Captured by the tails of the distributions reported in Fig. 5.7(A), a small fraction of clusters behaves differently and reveals

an ERS for *get-ready*, which might be caused by remaining artifactual components.

As already stated for the examples given in Fig. 5.5, the event- and cluster-specific maximal envelope differences $\Delta\phi_{max,avg}(m, c_k)$ vary between and even within subjects as exemplarily reported in Fig. 5.7 (B)–(E). The markers related to a single subject in Fig. 5.7 are very close or even overlapping for some cases, which is expected due to finding identical clusters over the twelve clustering runs per subject.

For brevity, this analysis is only provided for the NTik-SPoC analysis on dataset (D1a) but can in principle also be applied to the MI decoding scenario.

5.3.1.3 *Characterizing clusters by their associated components*

From a domain expert's point of view, it is interesting to know, which components are assigned to clusters by the DBSCAN method. This information becomes accessible through the approach.

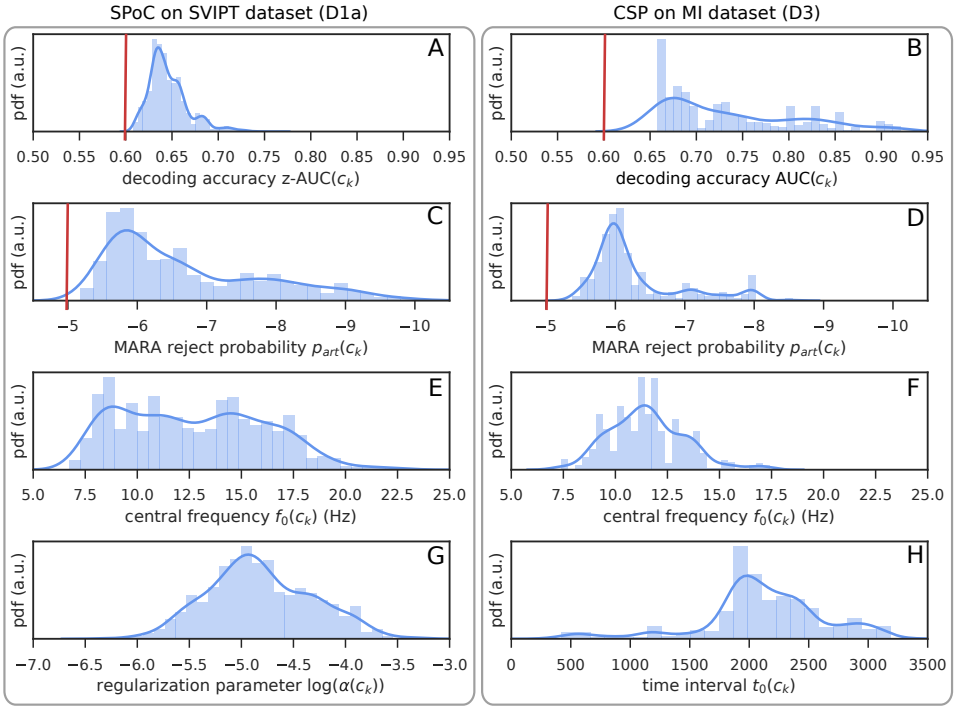


Figure 5.8: **Cluster characterization by within-cluster distribution of various parameters.** Distributions are contrasted for the two scenarios under investigation: (A) and (B) report the average within-cluster decoding accuracy, while (C) and (D) report the average MARA ratings. The vertical red lines refer to the denoising thresholds. (E) and (F) show the range of mean central frequency values. (G) reports the log-scaled average regularization parameter utilized for NTik-SPoC optimization, while (H) displays the average time parameter t_0 regarding the time interval for CSP filter training.

Table 5.1: **Group-level cluster statistics.** The numbers are gained across all subjects and twelve repetitions for kernel PCA preprocessing.

decoding scenario	dataset	selected config. $\lvert\Omega_{sel}\rvert/\lvert\Omega\rvert$	no. of clusters $\lvert C\rvert$
SPoC on SVIPT	(D1a)	$6.2 \pm 4.8\%$	7.4 ± 3.0
CSP on MI	(D2)	$10.5 \pm 3.2\%$	3.7 ± 1.7

In Fig. 5.8, the distributions for different quantities are shown which were obtained upon pooled clustering runs across all subjects for the two decoding scenarios.

The reliability of clustered components is shown in the first two rows of Fig. 5.8 with respect to their underlying decoding performance in (A) and (B), and resolved by their artifactual MARA rating in (C) and (D). The denoising step (see Sec. 5.2.3.2) required all components entering the clustering to provide a minimum decoding accuracy as well as to reveal minimal artifactual contamination. The plots nicely demonstrate that the majority of clusters clearly exceed these reliability criteria. It can be observed, that the MI decoding scenario reveals much larger decoding accuracies compared to the SVIPT scenario.

(E) and (F) show the average within-cluster central frequency. Combining this with the insight that within-cluster variations with respect to the central frequency parameter f_0 are rather small (see Fig. 5.9 (C) and 5.10 (C) later on), the distributions in (E) and (F) report the probability of finding clusters in specific frequency ranges. While for SVIPT the distribution is mostly dominated by alpha- and beta-band frequencies, the MI scenario mostly reveals clustered components in the alpha range.

Two distributions were specifically selected to characterize the underlying decoding scenario: in (G), the log-scaled distribution of the within-cluster regularization parameter utilized for the NTik-SPoC filter optimization is displayed. Its distribution is in good accordance with parameter ranges that were reported as suitable in Sec 4.4.1.5 and 4.4.2.2. (H) shows the average time point t_0 for the CSP training interval. On average, most clusters are found for CSP components trained on data aligned to an interval $2\,s$ after the *cue*.

5.3.1.4 *Group-level cluster statistics*

The cluster statistics on the group-level for both decoding scenarios is summarized in Tab. 5.1. For both decoding scenarios, we observed substantial differences across subjects regarding the relative number of available robust oscillatory components that survived the denoising step (see Sec. 5.2.3). Comparing both scenarios, for SVIPT the number of robust components after denoising is almost halved as in the MI scenario which emphasizes the higher overall complexity of the decoding task—as mentioned earlier in Fig. 5.8 (A) and (B). The overall number of identified clusters also differs substantially for both scenarios: in SVIPT, the number of identified clusters

is approximately doubled compared to the MI scenario which underlines the richness of the evoked within-trial envelope dynamics.

5.3.2 *Group-level validation of the approach*

For the validation of our proposed envelope mining approach, the final clustering results are analyzed on a group level in three manners: first, the homogeneity of the clusters are investigated by multiple validation metrics for both data scenarios including a comparison of different preprocessing methods. Second, the role of the density-based clustering compared to a standard k-means approach is investigated. Third, the comparison of different validation metrics reveals information on the quality of clustering. The following group level evaluation is based on pooled results across all subjects per dataset and the corresponding twelve clustering repetitions.

5.3.2.1 *Homogeneity of clusters*

Decoding scenario (1): NTik-SPoC analysis on SVIPT data

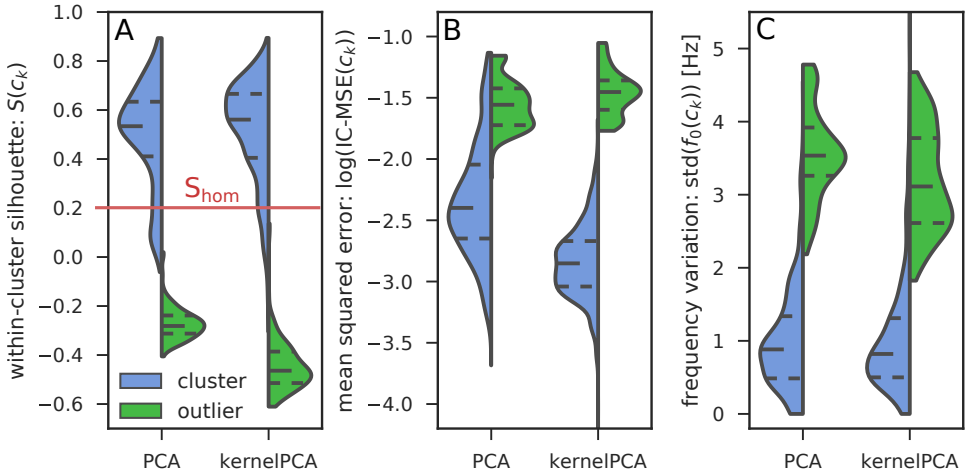


Figure 5.9: **Contrasting distributions of within-cluster evaluation metrics for two different preprocessing methods.** Based on a clusterings on SPoC envelope dynamics for SVIPT dataset (D1a), the results emerging from PCA preprocessing are contrasted to the ones of kernel PCA. Internal validation by average silhouette $S(c_k)$ is shown in (A). The horizontal red line corresponds to the global silhouette threshold S_{hom}. As external validation, the log-scaled IC-MSE (B) is provided. In (C), the within-cluster central frequency variation std $\left(f_0(c_k)\right)$ is shown. For each violin plot, the dashed horizontal lines refer to the two quartiles and the median.

DBSCAN separates between cluster and outlier samples. In Fig. 5.9, the distributions for three types of metrics (see Sec. 5.2.3.5) are contrasted for all identified clusters against detected outlier sets. In addition, the metrics are compared between the two applied dimensionality reduction methods. (A) depicts the distribution for the within-cluster silhouette $S(c_k)$.

As enforced by the applied selection criteria for ϵ^*, the distribution of $S(c_k)$ across all clusters is mostly above the subject-independent threshold criteria $S_{hom} = 0.2$, while the outlier distribution (blue) is mostly negative. In (B), the distribution of the within-cluster IC-MSE is shown. In accordance with the silhouette distribution in (A) but in reverse direction, the IC-MSE distribution is strictly shifted towards lower values compared to the outlier class. (C) reports the variation of the central frequency std $\left(f_0(c_k)\right)$ which is consistently below 3 Hz for clusters (mostly in the alpha- and beta-band range) and conversely not that confined for outliers. When comparing the two preprocessing methods across the subplots (A)-(C), we observed that kernel PCA results in more favorable partitionings than PCA. This is documented by, e.g., larger silhouette values in (A) or smaller IC-MSE values in (B). Unless noted otherwise, hereafter we report results based on kernel PCA.

Decoding scenario (2): CSP analysis on MI data

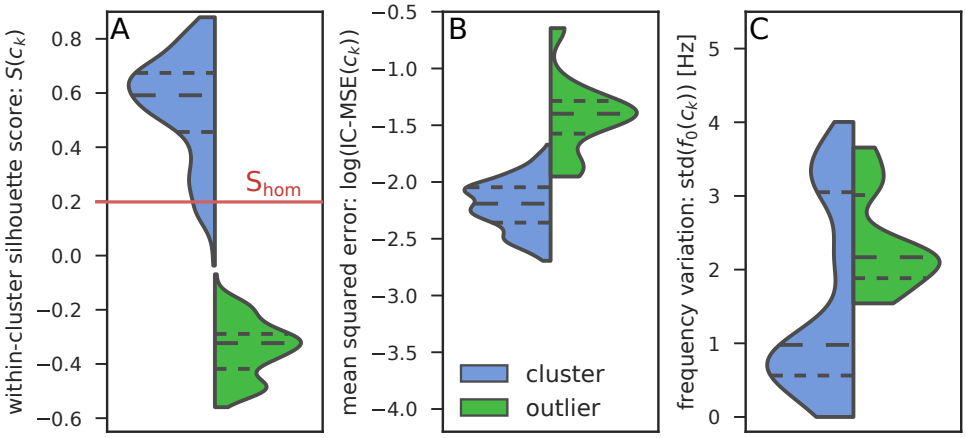


Figure 5.10: **Contrasting distributions of within-cluster evaluation metrics for clustering CSP components on the MI dataset.** In analogy to Fig. 5.9, (A) shows the average silhouette $S(c_k)$, while (B) provides the log-scaled IC-MSE and (C) displays the within-cluster central frequency variation std $\left(f_0(c_k)\right)$.

Based on clusterings of the second decoding scenario, Fig. 5.10 reports on the homogeneity of clusters in an equal manner as Fig. 5.9. Three main aspects can be stated: Fig. 5.10 (A) manifests that the vast majority of identified clusters are above the silhouette threshold $S_{hom} = 0.2$. In (B), a distinct separation between the cluster and outlier clusters is observable for our external validation metric while (C) shows that most clusters are again frequency confined below 2 Hz. These three observations are in accordance with scenario (1) and strongly support the general validity of harnessing envelope dynamics of spatial filter models. However, when comparing the absolute homogeneity of clusters measured by the IC-MSE metric, the MI scenario delivers not as homogeneous clusters as the SVIPT scenario. Furthermore, there are a few residual clusters which are rather large in size (data not shown) and also reveal a rather broad frequency variation,

captured by the cluster tail in Fig. 5.10 (C). These two observations for the MI scenario might be influenced by the lower signal-to-noise ratio of the envelope traces as for the MI scenario 75 trials per class and session were used compared to 400 trials for SVIPT. Furthermore, the number of identified clusters in the MI scenario is halved compared to the SVIPT scenario (see Tab. 5.1) which can also promote more inhomogeneous clusters.

5.3.2.2 *Comparison of density-based clustering with k-means*

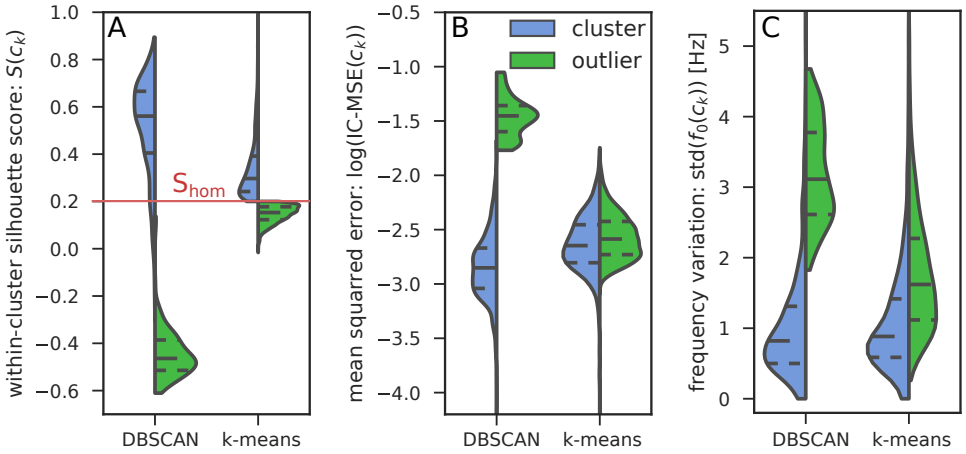


Figure 5.11: **Contrasting performance of DBSCAN with k-means.** The comparison is based on data of NTik-SPoC analysis for decoding scenario (1). In analogy to Fig. 5.9, the three established validation metrics — silhouette score (A), mean squared error (B) and frequency variation (C) — are shown.

As a baseline for the DBSCAN algorithm, it was substituted by k-means for the final clustering step (see Fig. 5.3) while keeping all other building blocks identical. Fig. 5.11 contrasts the performance of DBSCAN with k-means. While the reported DBSCAN results can be separated between cluster and outlier samples, k-means does not explicitly provide such an outlier notation. To compensate for this feature of DBSCAN, each cluster of a k-means run was post-hoc labeled as an outlier if the cluster homogeneity measured by the within-cluster silhouette $S(c_k)$ was below the subject-independent threshold. This resulted in a split between cluster and outlier samples for DBSCAN and k-means as shown in Fig. 5.11.

In Fig. 5.11 (A), the distribution for the within-cluster silhouette $S(c_k)$ is reported. This metric is finally optimized by the applied selection criteria for ϵ^*. While the silhouette distribution for DBSCAN is mostly above the subject-independent threshold criteria $S_{hom} = 0.2$, for k-means barely half of the distribution exceeds this value. This observation also translates to the external validation metric, the within-cluster IC-MSE (B), as DBSCAN reveals more homogeneous clusters compared to k-means. The only metric in which k-means is almost comparable to DBSCAN is the within-cluster central frequency variation. Here, both distributions are of similar shape, except that the tail of DBSCAN ends for lower frequency values. In summary,

for the shown data on scenario (1) we can report that DBSCAN reveals substantially more homogeneous clusters and is thus the favorable method if the identification of maximally homogeneous clusters is the ultimate goal.

5.3.2.3 *Comparing cluster (validation) metrics*

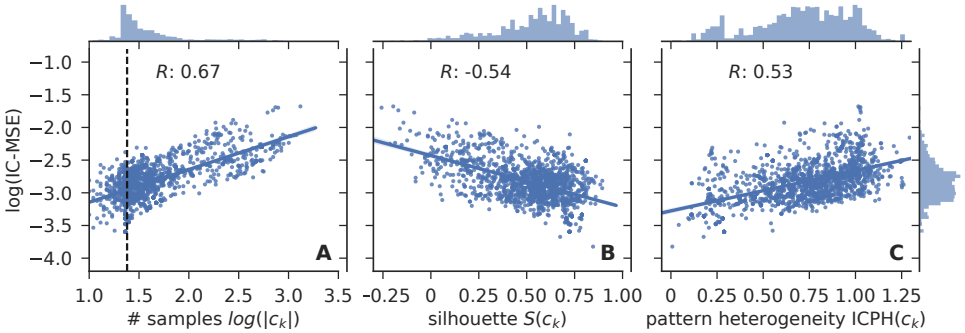


Figure 5.12: **Group-level analysis of cluster (validation) metrics.** On data of scenario (1), each point refers to one single cluster resulting of all performed clustering runs across subjects. Outliers are not shown here. For all subplots, the log-scaled IC-MSE is shown on the y-axis. Three different interactions are shown and the corresponding correlation is reported: (A) is referenced against the log-scaled cluster size. The dashed line displays the DBSCAN parameter $m_{pts} = 24$. In (B), IC-MSE is referenced against the within-cluster silhouette. (C) shows the interplay with the pattern heterogeneity. For each of the four metrics, the corresponding distributions are reported. © 2019 IEEE

The interaction between the external validation metric IC-MSE to the cluster size $|c_k|$ and two other evaluation metrics (see Sec. 5.2.3.5) is reported in Fig. 5.12. In (A), the averaged envelope homogeneity revealed by the IC-MSE is strongly dependent on the number of samples $|c_k|$ of the respective cluster, even though it corrects for the cluster size. The smaller the cluster size, the more probable it is to find a homogeneous cluster. The dashed line in (A) reports the DBSCAN parameter $m_{pts} = 24$. If border points are within ϵ-distance of core points of multiple clusters, they will be assigned to only one of them [208]. Thus, we find cluster sizes smaller than m_{pts}. Plot (B) shows a negative correlation of the external IC-MSE with the internal silhouette score which was used for DBSCAN optimization. The larger the silhouette value, the smaller the IC-MSE. Similarly in (C), there is a comparable positive correlation between the IC-MSE and the pattern heterogeneity. For small IC-MSE values, a high pattern homogeneity (reflected by low values) is found. However, we also found cases with low IC-MSE but substantial pattern heterogeneity. Highly similar interactions between the validation metrics were found for the MI decoding scenario (data not shown).

5.4 DISCUSSION

In summary, a method for an informed component selection which addresses and exploits the variability of spatial filter estimates was established. The method identifies groups of oscillatory components that satisfy two optimization criteria: (a) components are required to display a robust decoding performance, and (b) they need to reflect functional relevance for the given experimental task.

Applying these criteria, we demonstrated that the event-related envelope dynamics of oscillatory components can provide a rich source of information, as — in our two chosen data scenarios — these dynamics are strictly time-locked to within-trial events of underlying behavioral tasks. In addition, we showed how this information can be exploited to identify *reliable and functionally relevant* oscillatory features from a large hyperparameter space. Thereby, our data-driven approach is capable to deal with the noisy character of EEG data. Moreover, it was necessary to design the approach such that it can explicitly cope with rank instabilities as these typically are observed when dealing with eigenvalue decompositions on real-world datasets (see Sec. 3.4.4). These instabilities are commonly caused by slight variations of the training dataset, or when non-deterministic decoding approaches are utilized.

5.4.1 *Choice of features for clustering*

The clustering step allows assessing the reliability as well as the functional role of oscillatory brain signals. As discussed for the examples given in Fig. 5.5 and 5.6, exclusively utilizing the within-trial event-related envelope dynamics for clustering and no other features, such as scalp patterns, turned out to be a suitable choice for our data scenario. In Fig. 5.12 (C), we reported that single clusters can contain rather heterogeneous spatial activity patterns. This could mean that the neural origins of these oscillatory components differ despite the similarity in ERD/ERS features [59]. Accordingly, Bigdely-Shamlo et al. [202] showed that it might be beneficial to additionally consider 3D dipole locations for clustering, which result from a source reconstruction step. We agree on this view, as specifically from a clinical perspective knowledge about functional brain regions could provide added value [218]. However, source reconstruction comes at a price, as results are sensitive to initial assumptions and raw signal quality, among others [98, 219]. As additional source features would have enlarged our dimensionality, here we decided against including them into our feature vectors. Moreover, the results by Onton and Makeig [59] support the view that solely event-related dynamics might be sufficient to assess the functional role of oscillatory EEG features. Our observations support this judgment in the context of the investigated behavioral tasks.

Notably, our results show that the discovered clusters comprise components with highly similar, strictly confined frequency ranges. Overall, most clusters were found to represent oscillations of the alpha- and beta

range (see Fig. 5.8 (C) and (D)). Regarding the MI scenario, this finding is in accordance with literature as alpha- and beta-band features are commonly used for classification of left vs. right hand trials [220]. However, beta-band clusters were sparsely present in the MI scenario compared to the SVIPT scenario. A possible explanation might be the lower SNR of session-average envelope features for the MI dataset (only 75 trials per class). This limitation might aggravate the stable formation of beta-band clusters which intrinsically come with weaker ERD/ERS modulations due to the $1/f$ spectral power decrease in EEG signals (see Sec. 2.1.2). As such, measurable envelope modulations build the central prerequisite of the proposed mining approach. This argument might also explain that we did not identify clusters in the gamma-band range, even though we initially identified predictive components for the SVIPT decoding scenario (see Fig. 3.10).

5.4.2 *Design choices for the clustering step*

The clustering step required a limitation of the input dimensionality of the envelope features [206]. In this context, we contrasted the performance of PCA and kernel PCA for reducing the feature space. We could show that clusters detected after condensation with kernel PCA were found to be more *homogeneous* than those resulting from PCA preprocessing (see Fig. 5.9).

While the literature on clustering of brain signal features is mostly dominated by the k-means approach [201, 203], we chose the density-based DBSCAN algorithm and reported for the SVIPT scenario that it delivers favorable clusterings compared to k-means. Overall, DBSCAN holds various advantages for our clustering problem: first, DBSCAN does not assign each data sample to a cluster. Beside dense clusters, it handles outlier samples without assumptions about the global distribution of outliers. This is beneficial for our data scenario, as we can not expect every configuration to display well-defined envelope dynamics. Second, as the non-parametric DBSCAN does not make an explicit assumption about cluster shapes, it copes well with non-linearly separable clusters. In contrast, the parametric k-means method is biased towards convex cluster shapes and may not deal well with non-convex shapes. Motivated by similar arguments, Bigdely-Shamlo et al. [202] utilized an affinity propagation clustering to circumvent the aforementioned shortcomings of k-means.

Revisiting our complete approach, we find several hyperparameters which influence the outcome of the final clustering step. Among others, these comprise the number and width of the time intervals for the temporal subsampling (see Fig. 5.4) as well as the number of components D_{red} obtained from the dimensionality reduction step. The choice of these hyperparameters reveals a trade-off between an adequate (high) temporal resolution of the envelope dynamics and avoiding the curse of dimensionality for large feature dimensions [221]. Similarly, the log-scaling of the features could have also been renounced or replaced by other mapping

functions. Thus, in principle, the framework could be further optimized by, e.g., maximizing the correlation between within-cluster silhouette scores and the IC-MSE scores. However, this optimization remains outside the scope of this thesis.

5.4.3 *Method applicable for within- and across-subject clusterings*

To allow for a group-level analysis as well as for comparisons between subjects and single subject analyses, we decided to create equally sized datasets before running DBSCAN. For all but three subjects, this requirement lead to a downsampling of components, and we are aware, that we may have omitted informative data especially for subjects with a large component space Ω_{sel}. The repeated downsampling (an ordered random subset selection to Ω_{sel}) typically resulted in quite disjoint subsets, and each of them translated into different clusterings.

While we restricted the analysis to within-subject clustering, the proposed approach can also be applied to perform across-subject clusterings in the future [203]. However, this scenario is more challenging due to strong subject-to-subject variations in brain activity caused by, e.g., anatomical differences which evoke different scalp projections [90, 202] of the same functional sources.

5.4.4 *Many complex tasks provide a rich inner structure*

One key ingredient of the introduced method is the exploitation of the rich within-trial event structure of complex tasks, such as SVIPT (see Fig. 5.1). The event-related envelope dynamics of single oscillatory components enabled to partition a large number of oscillatory features into groups of functionally relevant and reliable components.

Comparing the two investigated datasets, the overall number of identified clusters was doubled for the SVIPT paradigm compared to the less complex MI paradigm. This observation seems plausible as a more complex within-trial structure could potentially evoke an enriched ERD/ERS dynamics resulting in a larger cluster count.

The within-trial structure is defined by the sequence of events that occur along a single trial of a paradigm. For SVIPT, these multiple events offer a large number of choices when it comes to extracting ERD/ERS features: oscillatory activity could have been extracted not only for hit events, but in addition also aligned to error events (such as when the cursor overshoots a target field) or other application-specific events. The proposed analysis concept should generally be applicable to many complex real-world tasks, as long as they reveal sub steps that can be utilized to define a "within-task" event structure. For some real-world tasks the definition of "within-task" structure might not be as straightforward or temporal markers of subtasks may not have been recorded. In these cases we propose to consider implicit behavior such as eye movements to identify sub-steps of an underlying

task. Exploiting this additional information, current brain state decoding approaches might gain access to the underlying functional role of features.

5.4.5 *Identifying functional roles of components on novel data*

Assuming we apply our method on novel data, we propose a simple two-step procedure for identifying functionally relevant components: first, the *reliability* of the identified clusters needs to be verified, e.g., using the external IC-MSE metric. A comparison to the distribution reported in Fig. 5.12 will enable to judge its reliability. Second, one can assess the functional contribution of a cluster's components by investigating the *ERD/ERS characteristics*. Functionally relevant components should reveal amplitude modulations time-locked to at least a few within-trial events (see Fig. 5.5 and Fig. 5.7). Along this line, the exact timing of ERD/ERS effects might provide valuable additional information.

5.4.6 *Expected benefit for targeted closed-loop interaction*

The offline results demonstrated the feasibility of the proposed approach even in challenging data regimes of chronic stroke patients. Thus, our method can provide a valuable offline tool to prepare informed closed-loop interaction protocols (see next Chap. 6). To state an example, we foresee a potential benefit in the field of BCI protocols for post-stroke rehabilitation training [12], in which our method could be beneficial in multiple ways.

First, it can allow gaining introspection about the training progress by monitoring underlying cortical processes across multiple sessions. As an example, the introspective character of our method may allow contributing to the current debate on the choice of signals to exploit for decoding the movement-related information. Current approaches use neural activity from the lesioned sensorimotor areas, from contralesional areas, a combination or any informative channels irrespectively from the lesion location [17, 222]. The monitoring of these various features over the course of a BCI-supported training now becomes accessible by our method.

Second, it may help to increase the efficiency of current BCI systems as it allows the practitioner to realize an informed feature selection such that functionally specific BCI feedback can be realized.

5.5 LESSONS LEARNED

We presented a data-driven method for assessing reliable and functionally relevant oscillatory EEG components estimated by different spatial filtering approaches. For this purpose, we first embrace the large variability of the generated spatial filters before condensing their functional signatures by the density-based DBSCAN clustering. For the first time, DBSCAN was successfully applied on brain signals and in contrast to k-means clustering,

this method allowed for the inclusion of an outlier cluster which proved useful as brain activity recordings typically provide a low SNR ratio.

As a novelty, we make use of within-trial structure for the condensing step. The approach was evaluated on data of two different motor task scenarios of diverging complexity: (1) on motor imagery data together with the CSP approach to identify class-discriminative oscillatory features. (2) On SVIPT data in combination with the NTik-SPoC algorithm to predict upcoming motor performance. Interestingly, in both scenarios we could show that the within-trial task structure translates into clusters of rich temporal dynamics of oscillatory components.

The proposed approach can be applied in combination with any spatial filtering algorithm given that the paradigm provides within-trial task structure. Providing introspection about individual, task-related ERD/ERS envelope signatures, we see the method's potential for understanding the neurophysiological and thus functional roles of components. Finally, an informed component selection may increase the efficiency of closed-loop protocols with feedback based on oscillatory activity, such as deployed in stroke rehabilitation.

6

MANIPULATING MOTOR PERFORMANCE BY CLOSED-LOOP BRAIN STATE INTERACTION

The conducted research of this chapter has not been published yet. Moreover, it is the starting point of the novel DFG project "SuitAble" for whose successful grant proposal I have substantially contributed.

SUMMARY

Repetitive motor tasks reveal strong trial-by-trial performance variations especially when dealing with subjects suffering from motor deficits. Based on a data-driven framework, we trained robust models to extract brain states from oscillatory brain activity which are predictive for the upcoming motor performance on single-trial level. The real-time estimation of these pre-trial brain states allowed the application of a gating strategy to specifically select *suitable* and *unsuitable* starting time points for a repetitive hand motor task. In a pilot study with four chronic stroke patients, we performed an intensive motor training with approximately 15 hours of effective training in eight online sessions. Even under challenging conditions with patients, we confirmed the expected hypothesis that in all four patients single-trial reaction times were overall significantly reduced for *suitable* trials. Moreover, shorter trial times under suitable states were found in two patients. Added value is provided by the introduced decoding framework as it allowed tracking brain signatures along the training and thus enables for a monitoring of training-induced changes.

A future randomized controlled study could investigate whether training under specific pre-trial brain states is beneficial for post-stroke motor learning. Overall, this successful proof-of-concept motivates to transfer this framework to other application fields, such as cognitive rehabilitation, sport sciences or systems neuroscience.

6.1 INTRODUCTION

Machine learning methods allow for the single-trial decoding from brain activity recordings, such as EEG, for real-time applications [19]. Particularly, it has frequently been demonstrated that a *direct* decoding of a user's intention from EEG signals can be achieved, such as the distinction between left and right hand movements [53] or target and non-target stimuli [223]. Recently, BCI systems were suggested to be also capable to extract and

exploit additional information about a user's current brain state by tracking corresponding informative neural processes [20, 21, 224].

Focusing on the field of post-stroke motor rehabilitation, a variety of BCI systems have been proposed and their efficacy—as well as the efficiency compared to non BCI-supported baseline methods—is still under intense investigation [12, 13, 17, 67]. In most applications, the BCI system exploits brain signatures, which are directly informative about an attempted, executed or imagined movement and subsequently triggers, e.g., a muscular stimulation, an orthotic device or sensory feedback (see Sec. 2.2.2).

A use case in which the inclusion of the user's current brain state might provide added value can be stated for studying repeated motor tasks as deployed in post-stroke rehabilitation. Here, behavioral intra-session performance variations are observed on two different time scales [28]. While intra-session trends mostly reflect motor skill acquisition, we showed that trial-by-trial performance variations on the scale of seconds can partially be explained by pre-trial oscillatory activity (see Chap. 3). Such oscillatory power fluctuations were found to be affected by the interaction of various networks—visual, premotor and motor cortex as well as subcortical and spinal structures [46, 123, 225].

Our findings now allow to exploit such predictive oscillatory brain states in real-time applications, which have rarely been transferred to closed-loop experiments. One of the few online studies recently showed that the targeted pre-trial modulation of sensorimotor rhythms can have an influence on upcoming task performance [226, 227]. However, a successful influence was only found in three of eight chronic stroke patients. While prospectively such systems might enable to augment post-stroke motor learning, such complementary BCI systems need to ensure a *reliable* decoding of *functionally relevant* features under challenging conditions with patients.

Here, the contribution of this thesis chapter comes into play: we now utilize the established workflow for single-trial motor performance prediction based on the data-driven extraction of informative oscillatory brain signal components (see Chaps. 3 and 4) and the subsequent contribution to identify functionally relevant spatial filter models (see Chap. 5). Thereby, these earlier findings are validated in an online scenario. We hypothesize, that the computed performance prediction models directly enable to extract individual predictive brain states which allow to influence upcoming single-trial motor performance. Specifically, the pre-go power of a robust, predictive component shall be evaluated in real-time to specifically select between *suitable* and *unsuitable* starting time points for SVIPT. Exemplified for a multi-session post-stroke hand motor training, the hypothesis will be evaluated in a pilot study.

Research question Q3

6.2 METHODS

6.2.1 *Subjects*

Tab. 6.1 reports on the demographic and impairment related data of the four included chronic stroke patients abbreviated by *P*1 to *P*4. Here, the term *chronic* implies the stroke to be dated back at least 3 months before participation [228]. The patients were selected to reveal their first-ever, unilateral ischemic stroke resulting in a mild to moderate hemiparesis (except *P*1 revealed a stronger impairment). Three of four patients were naïve to SVIPT and BCI applications, respectively. *P*4 had prior SVIPT experience due to participation in the control group [228], *P*2 completed a 30 hour language training with feedback based on task-relevant EEG activity [70].

Prior to participation, all subjects provided written informed consent. The study was approved by the local ethics committee of the University Medical Center Freiburg.

Table 6.1: **Patient characteristics of dataset (D1c).** Demographic and impairment related data of the four chronic stroke patients included for the pilot study. The listed patients are disjunct with the ones of dataset (D1b).

Patient	*P*1	*P*2	*P*3	*P*4
Age (years)	44	64	56	52
Gender (male/female)	m	m	f	m
Affected limb (right/left)	r	r	r	l
Time after stroke (months)	37	83	24	59
Initial Upper Extremity Fugl-Meyer (UEFM)	27	52	53	58
Naïve SVIPT user (yes/no)	y	y	y	n
Naïve BCI user (yes/no)	n	y	n	n

6.2.2 *Experimental setup*

All four pilot patients completed a high-intensity hand motor training with their affected hand comprising about 15 hours of effective training time. As shown in Fig. 6.1(A), the full training was overall composed of 10 training sessions—thereof 8 online sessions—of the EEG-tracked SVIPT within maximal 3 consecutive weeks[1].

Before the first online session and after the last training session, a clinical assessment of the patients was performed. It contained the evaluation of the patient's maximal pinch and grip force, as well as the widely established Upper Extremity Fugl-Meyer (UEFM) score [229, 230].

A single SVIPT training session consisted of ten runs with 20 trials each. Throughout each session, the patient's EEG signals were registered from 63

1 Patient *P*1 performed *three* offline and eight online sessions. As *P*1 had to cope with muscular fatigue in six sessions, in those ones only 5-9 runs were performed.

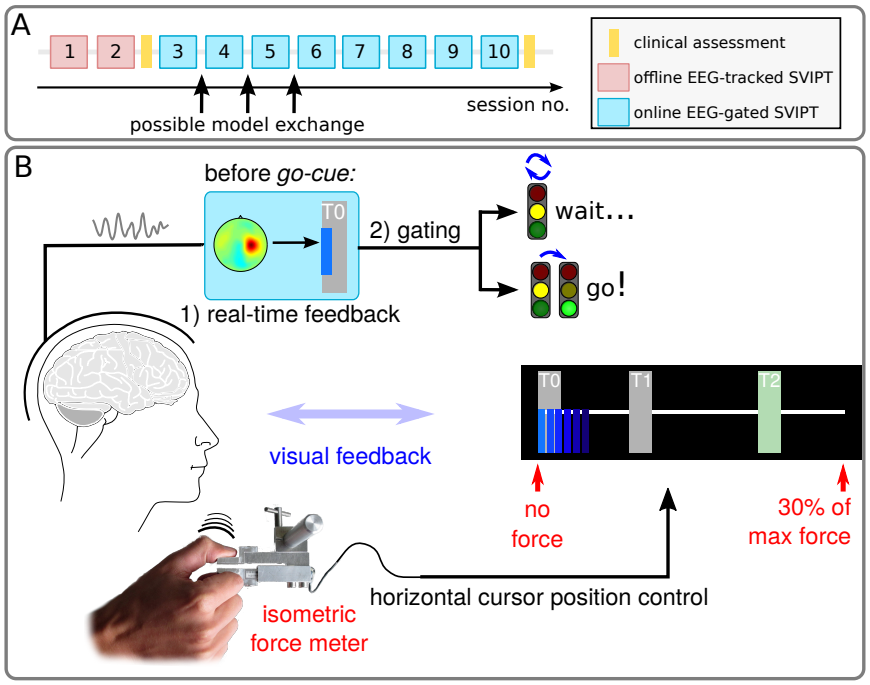


Figure 6.1: **Experimental protocol of the online EEG-gated SVIPT.** (A) Study protocol for testing motor performance separability across eight online sessions. Before and after the online training, a clinical assessment was performed. (B) Scheme for the online EEG-gated SVIPT. Prior to the start of each trial, the patient received continuous feedback about his current brain state visualized by a real-time adaptation of the vertical cursor position. A state-dependent gating strategy determined the *go-cue* time point of every single trial. For details on the gating strategy, see Fig. 6.2.

passive Ag/AgCl electrodes placed according to the extended 10-20 system. If not explicitly noted differently, further experimental details of a single EEG-tracked SVIPT session were identical to the performed offline studies described in Sec. 3.2. After each run, an individual highscore with the best average trial durations per run was displayed to motivate the patients to further improve their SVIPT motor performance.

The offline sessions were used for the calibration of a performance prediction model, which is further described in Sec. 6.2.2.1. In each online session, the system tried to influence the patients' upcoming performance in the *get-ready* phase by: (1) providing continuous feedback about the ongoing brain state to the patient as sketched in Fig. 6.1(B). This was realized by a continuous adaptation of the vertical cursor position. As a reminder, during the SVIPT *get-ready* phase a slight blue cursor is presented in leftmost target field *T*0 while subjects are asked to fixate on the center of *T*0. (2) Depending on the current brain state estimate, a temporal gating strategy was realized prior to the start of a trial. A *go-cue* was elicited either if a user-specific prediction model indicated a desired brain state or if a timeout criterion was met (for details see Sec. 6.2.2.2).

In the offline sessions, the patients unknowingly received pseudo-feedback regarding their ongoing brain state as the vertical cursor position was altered according to earlier recorded brain state estimates of another pilot subject.

6.2.2.1 *Performance prediction models*

The calibration procedure to compute a performance prediction model for an individual patient combined all developed buildings blocks of this thesis. As a reminder, the detailed workflow for predicting single-trial motor performance from oscillatory EEG data was described in Sec. 3.2.5.

Based on the findings in Chap. 4, we chose the NTik-SPoC variant with Tikhonov regularization for this patient scenario with a strictly limited amount of training data. The algorithm was trained on pooled patient-specific data across the offline sessions, while the first 80 % (in chronological order) of the data from each session were used. The remaining session data were used for validation of the obtained models.[2] Since reaction time (RT) revealed the largest decoding accuracies in the earlier reported offline analyses (see Sec. 3.3.5), we selected this most promising metric as label input for NTik-SPoC and thus aimed to primarily influence upcoming RT in the online SVIPT sessions. In analogy to Sec. 5.2.2.1, we again embraced spatial filter variability under different hyperparameter configurations by training NTik-SPoC components in a large configuration space. According to the above described validation procedure, the cross-validation hyperparameter was dropped here. Based on the component's envelope dynamics across all offline trials, the component mining framework (see Sec. 5.2.3) was then also carried out to gain potential candidates for the closed-loop interaction.

In a final step, a manual inspection was performed among different feature candidates to select *one* specific oscillatory component $\mathbf{w}$ for the closed-loop interaction. Hereafter, this will be referred to as *selected* oscillatory component. According to the denoising step (see Sec. 5.2.3.2), all candidate components were required to provide a minimal predictive power z-AUC$_{min} \geq 0.6$ on the validation data and a MARA rating of their artifactual probability below $p_{art} \leq 10^{-5}$. In addition, the following selection criteria were taken into consideration:

1. **Rich envelope dynamics**: Driven by the findings of Chap. 5, oscillatory components revealing a rich within-trial envelope dynamics were preferred. As shown on SVIPT datasets (D1a) and (D1b), an ERD effect with *get-ready* and/or *go-cue* as well as an ERS with *hit 3* or *hit 4* substantiate the neurophysiological plausibility of individual components.

2. **Motor-related spatial activity patterns**: Regarding the spatial activity patterns of the components (see Eq. (2.3)), a high similarity (visually

2 For patient *P*1, the model was trained on the first two sessions and evaluated on the third one

inspected) across the offline sessions was required to ensure across-session stability of the target component. High priority was given to motor components, specifically lateralized over the patient's affected hemisphere.

After each of the first three online sessions, a validation on the unseen session data was done among all component candidates. In case the prediction performance of such a candidate component was clearly outperforming the selected one and simultaneously fulfilling all above stated criteria, the originally selected component/decoding model was then exchanged.

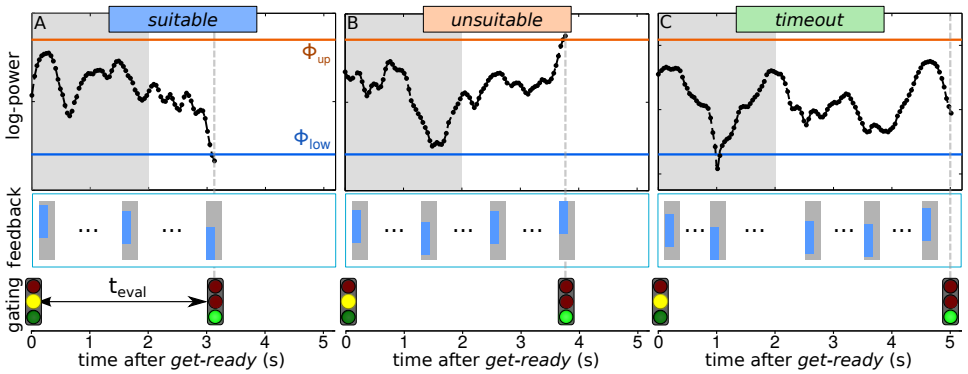


Figure 6.2: **Single-trial online gating for triggering a *go-cue*.** Arranged in the columns, example data for three single trials visualize the applied gating scheme during the *get-ready* phase. Rows show the continuosly sampled power, the displayed cursor feedback for the patient and the *go-cue* time point. All rows share the time axis. Initiated by the *get-ready* event (yellow light), the power of the selected oscillatory component was continuously estimated (black dots) and translated into a vertical cursor position. The *go-cue* time point (green light) was triggered earliest 2 s after *get-ready* and by distinguishing between three cases: (A) if the power $\Phi(t)$ fell below the threshold Φ_{low}, a *suitable* trial was elicited. (B) If $\Phi(t)$ exceeded the threshold Φ_{up}, an *unsuitable* trial was initiated. (C) A *timeout* trial was registered and still started, if the power did not exceed either of the thresholds in the interval $[2, 5]$ s.

6.2.2.2 *Brain state-dependent gating strategy*

The selected subject-specific spatial filter $\mathbf{w}$ was now utilized in a closed-loop setting that is sketched in Fig. 6.1 (B). During the *get-ready* phase (prior to the *go-cue*) of a single SVIPT trial, the component's log-bandpower $\Phi(t)$ was continuously evaluated every 40 ms according to the variance approximation given in Eq.(3.1). For brevity, power estimation hereafter always refers to log-bandpower estimation. Therefore, data of the latest 400 ms for alpha components (300 ms for beta components) sampled with 1 kHz was utilized. Those values were determined on earlier pilot subjects. The choice of the time window reveals a trade-off between the temporal resolution to capture optimal *go-cue* time points and sufficient data to estimate the ongoing power of the signal. The online data were bandpass

filtered by a linear Butterworth filter to the same frequency band as the training data of the selected model.

The closed-loop interaction was realized in two ways. First, the ongoing estimated brain state according to the component's power was directly translated into a vertical cursor position for the patient (see second row of Fig. 6.2). Second, a brain state-dependent single-trial online gating strategy was applied. Therefore, *go-cue* events were preferably triggered if according to the spatial filter model a highly preferable or extremely poor upcoming reaction time was expected. Given the current power estimate $\Phi(t)$, a *go-cue* event was triggered by a threshold scheme as sketched in Fig. 6.2: a *suitable* trial was elicited if the component power was registered below a threshold Φ_{low}, while an *unsuitable* trial was triggered if $\Phi(t)$ exceeded an upper threshold Φ_{up}.[3] If none of the thresholds were passed $5\,s$ after the *get-ready* event, the *go-cue* was immediately given and the trial labeled as *timeout*. The determination of both power thresholds will be addressed in the section below. Note, that in the offline sessions the time t_{eval} between *get-ready* and *go-cue* was sampled from the interval $[2, 5]\,s$ based on statistics of earlier pilot online sessions. For consistency, the same time limits were also applied throughout all online sessions. Hence, the earliest *go-cue* was possible $2\,s$ after *get-ready* as marked by the gray shaded intervals in Fig. 6.2. Threshold exceedances within the first two seconds, as sketched in Fig. 6.2 (C), were ignored.

In the online sessions, the patients were instructed to find a strategy to lower the vertical cursor position by modulating their brain signals during the *get-ready* phase. They were instructed that the cursor position was reflecting the quality of their upcoming motor performance. However, the subjects did not know that RT was the targeted motor performance metric.

As single sessions were strictly limited to only 200 trials (corresponding to $\approx$ 90 min), an online artifact detection was applied throughout the *get-ready* phase to reduce the rate of artifactual trials [42, 231]. Therefore, a min-max threshold of $100\,\mu V$ was applied on bandpass-filtered data (to the range $[0.7, 45]\,Hz$) of frontal EEG channels. In case of a threshold violation, the trial was immediately aborted and patients received visual feedback, followed by a restart of the trial after a $2\,s$ pause. Such aborted trials did not enter the post-hoc analysis.

6.2.2.3 *Online adaptation of the prediction model*

From a clinical point of view, we intended to control the ratio of *suitable* to *unsuitable* trials during training as we prospectively expect an influence of the ratio on post-stroke motor learning. Precisely, we decided to overall trigger $p^*_{sa} = 0.55$ *suitable* trials and $p^*_{us} = 0.35$ *unsuitable* trials to provide a fixed success rate to the patient while ensuring enough data for both conditions (required for statistical comparison). By that, we expected 10 %

3 This mapping of thresholds to labels assumes a negative correlation of single-trial power $\Phi(e)$ with single-trial motor performance $z(e)$. This relation can be determined by evaluating the sign of the correlation $R(z, \Phi(\mathbf{w})))$ on the training data.

timeout trials in total. The gate ratio was controlled by a slight online adaptation strategy as hereafter explained.

Given the calibrated model by the selected oscillatory component $\mathbf{w}$, when switching from the offline to the online phase the underlying component power will most probably follow different distributions [232]. This is commonly known as *covariate shift* [233]. In the field of BCI applications, various approaches have been proposed to account for the non-stationarity characteristics of neural signals in closed-loop applications [234–237]. In this pilot study, the selected individual filter model $\mathbf{w}$ was kept fixed and solely the decision boundaries, namely the gating thresholds Φ_{low} and Φ_{up}, were adapted along the online sessions.

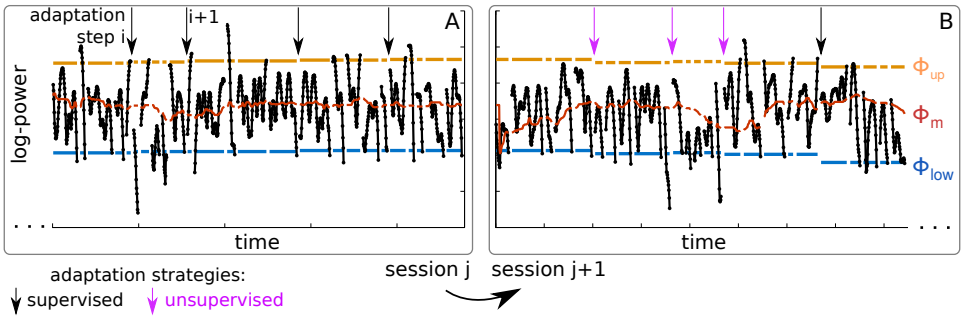


Figure 6.3: **Online adaptation strategies for the prediction model.** The continuously sampled pre-trial log-bandpower (black line) of a *selected* component is shown on exemplary data of *P*4. The data for each pre-trial phase—1.5 s after *get-ready* up to *go-cue*—is successively reported for the last (first) 25 trials from the end of session 9 in (A) (the start of session 10 in (B)), respectively. For clarity, after each trial—when the gating threshold was exceeded—a small gap is introduced. In (A), only refined supervised adaptation steps were performed every 5 trials to update the gating thresholds Φ_{up} and Φ_{low} by means of the current median estimate Φ_m. With the start of a new session (B), the associated unsupervised coarse adaptation steps are highlighted. Overall, the temporal distance of adaptations varies as the pre-trial phase is different from trial-to-trial.

To accomplish both requirements to reach an intended gate ratio while also coping with non-stationarities of brain signal recordings, two different strategies were deployed during online training as shown in Fig. 6.3:

- **(1) Coarse unsupervised adaptation**: To cope with non-stationarities in oscillatory fluctuations from session j to $j+1$ or after longer breaks within a session, a coarse unsupervised adaptation step was performed. We make the simplified assumption that the component's log-bandpower fluctuation width is stable across sessions while the average power level $\Phi_m(j)$ might be shifted from session j to $j+1$, as observed for CSP features [238]. Based on the median of the sampled power $\Phi_m(j+1)$ of the new session, the latest gating thresholds from the previous session j were updated by:

$$\Phi_{up}(j+1) = \Phi_{up}(j) + (\Phi_m(j+1) - \Phi_m(j)) \tag{6.1}$$

In analogy, $\Phi_{low}(j+1)$ was determined by adapting the last lower threshold $\Phi_{low}(j)$ of the previous session to the novel median power level. The respective power medians in Eq. (6.1) were estimated in a session-specific ring buffer containing the latest 150 sample points. As in each novel session $j+1$ the buffer needed to be filled up with data, this strategy was deployed throughout the first run of each online session every five trials as reported in Fig. 6.3. The buffer was only updated between 1.5 s after *get-ready* to the *go-cue*. An earlier interval would have introduced a systematic bias as *get-ready* triggered an ERD effect in most components (see Fig. A.3). In the first online session, the gating thresholds were initialized by the 5^{th} and 95^{th} percentiles of the power distribution of the selected component $\mathbf{w}$ on the initial training data.

- **(2) Refined supervised adaptation**: To reach the intended gating ratios, a refined, supervised adaptation from step i to $i+1$ was regularly applied every five trials during the online sessions, starting with the second run. To this end, the sampled probabilities p_{sa} and p_{us} of both gating conditions were evaluated on the latest 60 trials and used as labels to guide the adaptation. The applied iterative adaptation rule took two aspects into account: First, the adaptation was proportional to the signed quadratic condition-related probability deviation $\delta p_{us} = (p_{us} - p^*_{us})$ —and similarly for δp_{sa}—to penalize strong deviations from the expected label distribution. Second, the update was relative to the absolute distance of the previous threshold $\Phi_{up}(i)$ to the power median Φ_m based on the previously mentioned ring buffer of the current session $j+1$.

  This translated into an update scheme from adaptation step i to $i+1$ for the upper gating threshold:

$$\Phi_{up}(i+1) = \Phi_{up}(i) + \eta \cdot \text{sign}(\delta p_{us}) \cdot (\delta p_{us})^2 \cdot |\Phi_{up}(i) - \Phi_m| \quad (6.2)$$

  with a fixed learning rate $\eta = 2$ which was determined on earlier pilot data. In analogy, the lower gating threshold update $\Phi_{low}(i+1)$ can be estimated from Eq. (6.2) by inserting $\Phi_{low}(i)$ and δp_{sa} instead of $\Phi_{up}(i)$ and δp_{us}. To accomplish a careful adaption, the class-related probability deviations δp_{us} and δp_{sa} were required to exceed an absolute tolerance level of $\delta p_{tol} = 0.05$. Smaller condition-specific deviations did not result in a threshold update. After each run, the experimenter received feedback on the within-session component's power time course and the selected thresholds. In case of severe non-stationarities, e.g., when noisy EEG channels had to be fixed within a break, the experimenter was able to perform an optional unsupervised adaptation (see Eq. (6.1)) with the start of each run.

6.2.2.4 *Data preprocessing*

For post-hoc evaluation, the raw EEG signals were low-pass filtered at 100 Hz, sub-sampled to 500 Hz sample rate before high-pass filtering at

1 Hz. For frequency filtering, linear Butterworth filters of 5^{th} order were applied.

Noisy channels of single session recordings were removed by a two-step procedure. First, the variance of single epochs and channels was computed. Based on the pooled statistics, all cases outside the $[10, 90]$ percentiles and also exceeding twice the corresponding inter-percentile range were registered as outliers. Second, channels which allocated more than 10% of all outliers and a minimum of 5% outliers across epochs were removed.

Furthermore, artifact cleaning was done by an ICA decomposition on pooled data of the active trial phase—from *get-ready* to *trial end*—of each session. To restrict the computational effort for the ICA, a randomly selected run of each session was utilized.[4] The obtained ICs were rated for artifactual origin with the automated artifact detection framework MARA [89, 239]. Based on MARA ratings, a maximum number of 10 probably artifactual ICs were removed from the all single session EEG data before projecting it back into the original sensor space. Only this pre-cleaned data was used in the post-hoc feature analysis.

Based on the force sensor recordings, the different single-trial motor performance metrics were extracted for each single trial (see Sec. 3.2.3). In addition, the trial duration (TDUR) was computed. For the behavioral post-hoc analysis, the first 10 trials of each session were omitted due to the ramp-up phase of the threshold adaption with the start of each session.

6.3 RESULTS

The achieved results are organized as follows: First, we provide evidence that the brain state selection separates the gating conditions—*suitable* and *unsuitable*— as expected. Second, we present single-trial behavioral performance results contrasted for the gating conditions. Third, an in-depth feature introspection on the utilized oscillatory components for the closed-loop training is given. Fourth, the clinical assessment scores are shown to give an indication about individual post-stroke motor learning.

6.3.1 *Brain state-dependent gating*

As described in Sec. 6.2.2.2, the power of a selected oscillatory component was continuously sampled after *get-ready* to specifically identify *suitable* starting time points for SVIPT. As a reference for the behavioral validation, a smaller ratio of *unsuitable* trials was sampled. On data of each patient, Fig. 6.4 (A) reports the single-trial power distributions of the selected oscillatory component at *go-cue* on pooled data (after post-hoc artifact rejection) across all online sessions. Significance of the power contrast was tested by a two-sided Wilcoxon rank-sum test.

4 For *P*1, the sessions 1 and 2 were excluded for ICA training due to instabilities in the resulting decompositions.

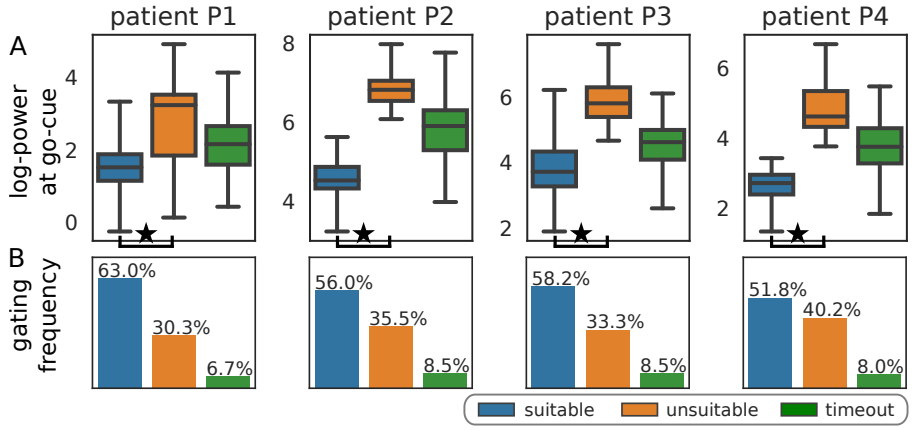


Figure 6.4: **Brain state separation at *go-cue*.** The shown results are obtained on pooled data across all online sessions and split by the three gating conditions. (A) The boxplot for each patient reports the component power that triggered a *go-cue* event contrasted for the different gating conditions. Statistical significance is indicated by * with $p < 0.01$. Each box shows the quartiles of the underlying data, the whiskers refer to 2·IQR. (B) The achieved frequencies of gating conditions across all sessions are displayed for each patient.

As intended by the single-trial gating strategy, there is a significant split between the *suitable* and *unsuitable* power distributions across all online sessions of all four patients. For $P2$ and $P4$, there is even no overlap of the boxplots observable which is due to stable across-session power fluctuations of the selected subspace components. Regarding the distribution observed for $P1$, the gating thresholds were adapted according to a percentile strategy of the recently sampled component power which led to just a temporal but not a global power separation. As another plausibility check of the established online framework, the *timeout* boxplot is located between the other two extreme distributions on the data of all four patients. Fig. 6.4 (B) reports the achieved ratios across the different gating conditions. As foreseen by the adaptation parameters p^*_{sa} and p^*_{us} (see Sec. 6.2.2.3), all four patients trained under the majority of *suitable* brain states. Furthermore, the gating statistics of $P2$ to $P4$ nicely demonstrate that the intended gating ratio was achieved within the given tolerance.

6.3.2 *Single-trial motor performance caused by different gating strategies*

It is now of specific interest to verify the single-trial motor performance that followed the previously detected brain state at *go-cue*. As the decoding model was trained on reaction time (RT), this metric will be reported first. In a second step, the transferability to other motor performance metrics that integrate behavioral information on longer trial intervals will be verified.

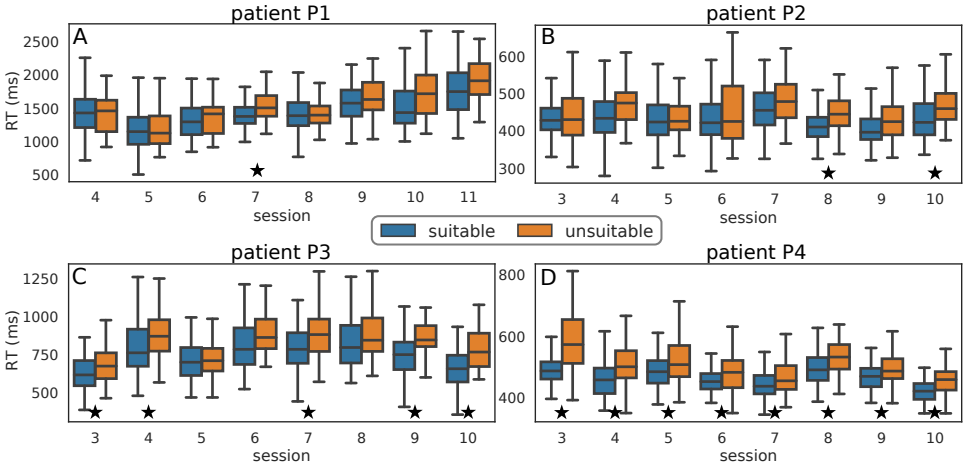


Figure 6.5: **Comparison of single-trial reaction times contrasted for the two gating strategies.** For the patients $P1$-$P4$, the RT distribution for *suitable* and *unsuitable* trials is shown across all single online sessions. A star refers to single sessions with a significant difference between conditions (significance level of $p < 0.05$ with Holm-Bonferroni correction for multiple testing).

Effect on single-trial reaction time (RT)

To test the central hypothesis of this chapter, namely, that upcoming motor performance can be influenced by the ongoing brain state of the patient, Fig. 6.5 reveals the behavioral results of the *go-cue* selection strategy. In (A)–(D), the RT distributions for each patient across every single online session and contrasted for the two different conditions are presented. The significance of the obtained contrast in single sessions was tested by a two-sided Wilcoxon rank-sum test. Even though two consecutive sessions are not expected to be independent, a conservative Holm-Bonferroni correction was applied on individual subjects.

In two of four pilot patients we found that single-session RT distributions separate significantly in at least half of the sessions, especially in the second half of the online training. Interestingly, in the best case of patient $P4$ a significant difference throughout all eight online sessions was achieved. Even though the online framework was tested with a heterogeneous patient group regarding their initial hand motor impairment (UEFM varied from 27-58 as shown in Tab. 6.3), on pooled data across all online sessions a significantly shorter RT for *suitable* trials can be stated for *all four* patients (see Fig. 6.6 (A)). Hence, the intended RT manipulation worked out successfully. Considering individually observed behavioral effects along the training, the median difference between the two opposing conditions relative to the interquartile range (IQR) varies between 27 % ($P2$) and 40 % ($P3$). The RT range of patient $P1$ reflects the severe motor impairment, which suggests handling the corresponding data with caution.

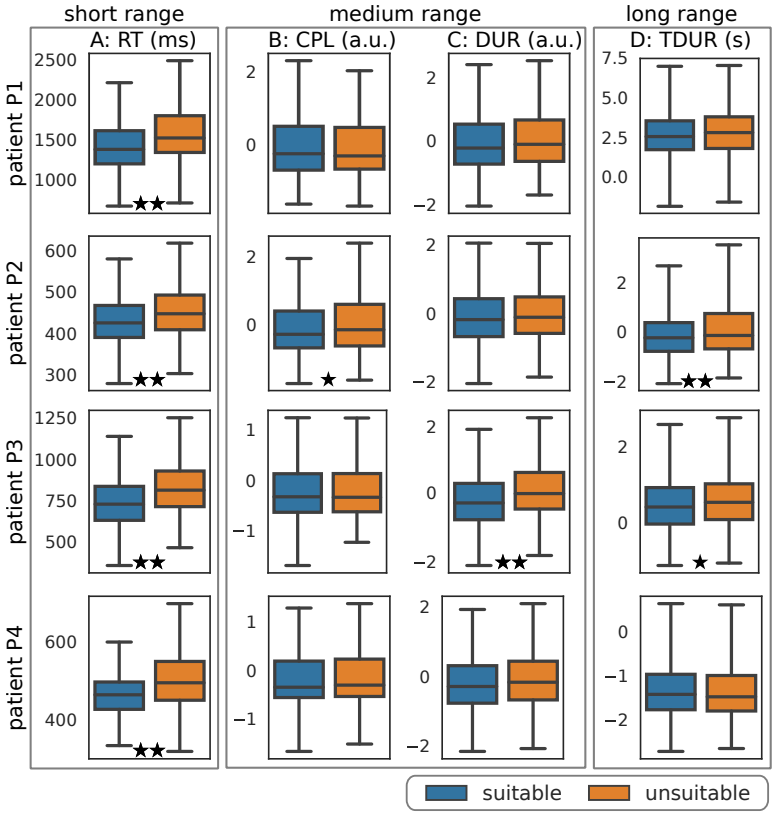


Figure 6.6: **Pooled single-trial motor performance data contrasted for the two gating strategies.** The boxplots on pooled data of different motor performance metrics column-wise aligned by their temporal integration range after *go-cue*: (A) reaction time, (B) cursor path length up to *hit 1*, (C) time to *hit 1* and (D) full trial duration (session-wise median subtracted). Each row refers to a single patient. The significance level for each tested metric (with Holm-Bonferroni correction) is reported by $p < 0.01$ (**) and $p < 0.05$ (*).

Indirect transfer to other motor performance metrics

While the utilized NTik-SPoC performance prediction model was optimized for RT, it is of special interest if preferential brain states at *go-cue* translated into improved motor performances captured by metrics integrating longer periods of the trial. On pooled data of all online sessions, Fig. 6.6 reports the individual distributions of different single-trial performance metrics. Their arrangement from (A)–(D) is orientated by an increasing temporal integration of behavior along the trial. While RT does only take a short time window after *go-cue* into account, cursor path length (CPL) and duration (DUR) were calculated up to *hit 1* ($\approx$ 40% of full trial time). Both metrics were standardized per session and subsequently pooled for both hit sequences. RT, CPL and DUR have been found to be rather uncorrelated among each other [30]. For the trial duration (TDUR), the session-wise median was subtracted before pooling to eliminate session-to-session differences (see Fig. A.5). Significance between conditions was tested by a

two-sided Wilcoxon rank-sum test with Holm-Bonferroni correction for testing multiple metrics.

While on pooled individual RT data a homogeneous picture in terms of a distinct separation was noticed in all four patients, the brain state-dependent gating does only partially translate on single-trial metrics which integrate performance over longer time intervals. Intriguingly, for half of the pilot patients a significant effect on SVIPT trial duration can be stated even though the utilized decoding model was not trained on these labels.

6.3.3 *Across-session feature introspection*

Hereafter, the presented results are based on a post-hoc analysis of the recorded EEG data across the single training sessions.

As the utilized spatial filter model to extract an oscillatory subspace is kept fixed during the online training, these features were required to be robust across sessions. In other words, each online session serves as an unseen validation dataset. Details about the selected oscillatory components are summarized in Tab. 6.2. Hereafter, various stability aspects regarding the utilized oscillatory subspace components are reported over the course of the training.

Table 6.2: **Details on the selected decoding model for closed-loop interaction.** For each patient, the hyperparameters as well as the initial decoding accuracy are reported for the *selected* spatial filter model. If a component was switched during online training, then two entries per patient are specified. In the first row, the respective sessions are reported.

patient	*P*1	*P*2	*P*3	*P*4
central frequency $f_0[Hz]$ (session)	9.5 (4) 8.7 (5-11)	21.8	18.8 (3-5) 10.9 (6-10)	9.7
reg. parameter α	$1.0 \cdot 10^{-2}$ $8.5 \cdot 10^{-4}$	$4.4 \cdot 10^{-5}$	$8.7 \cdot 10^{-6}$ $3.4 \cdot 10^{-3}$	$1.0 \cdot 10^{-2}$
training epochs N_{train}	342	344	289	347
decoding accuracy z-AUC	0.62 0.67	0.62	0.65 0.69	0.69
MARA reject prob. p_{rej}	$1.43 \cdot 10^{-9}$ $5.56 \cdot 10^{-9}$	$1.22 \cdot 10^{-6}$	$6.81 \cdot 10^{-6}$ $3.28 \cdot 10^{-7}$	$7.44 \cdot 10^{-9}$

Spectral and spatial feature analysis

The robustness of individually *selected* components—under scarce training data and across multiple sessions—was an important characteristic for this pilot study. The features' robustness was favored by the developed methods of Chaps. 4 and 5. For validation purposes, Fig. 6.7 displays the spectral content of the *selected* individual spatial filters $\mathbf{w}$ which were deployed for the online gating strategy. Therefore, the non-frequency filtered data

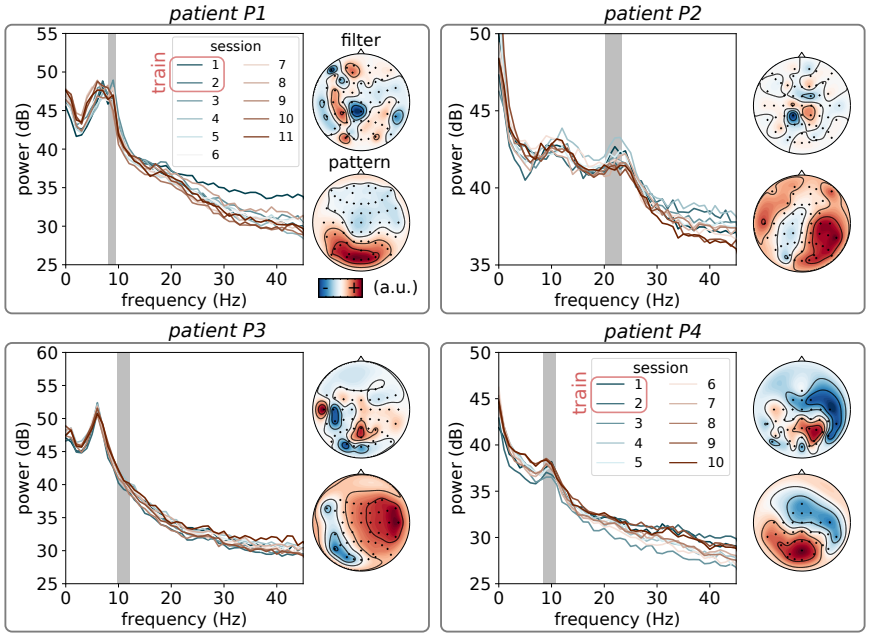


Figure 6.7: **Across-session spectral analysis of selected oscillatory components.** For each patient, the pre-go power spectra of the *selected* subspace components for the online gating are shown. The gray shaded area in each power spectrum refers to the frequency band in which the training data were filtered to train the model. Patients $P2$-$P4$ all share the same legend. In addition, the individually selected spatial filters and corresponding pre-go spatial activity patterns averaged across all sessions are displayed for each patient.

were projected to the *selected* subspace and segmented to the pre-go phase on data of the single sessions. The resulting session-averaged spectra are shown for each patient across the full hand motor training. For patients $P1$ and $P3$, the last *selected* oscillatory component is shown. On data of all four patients, the power spectra of the selected features are highly similar across sessions and reveal the well-known spectral $1/f$ decaying behavior with distinct α- and/or β-modulations. This observation is a strong indicator for the across-session robustness of the *selected* components.

An additional neurophysiological introspection is provided by assessing the spatial activation patterns of an oscillatory component (see Eq. (2.3)). As documented in the appended Fig. A.7, the selected oscillatory components revealed mostly stable pre-go activity patterns over the course of the multi-session training. This again supports the successful selection of robust oscillatory components.

Event-related envelope dynamics

As emphasized in Sec. 5.3.1.1, the SVIPT paradigm contains a rich within-trial structure which nicely translates into a distinct event-related envelope dynamics of certain oscillatory components. Exemplified for patient $P2$, Fig. 6.8 reports the induced envelope dynamics of the selected filter $\mathbf{w}$ extracted for every single session along the training. The average log-

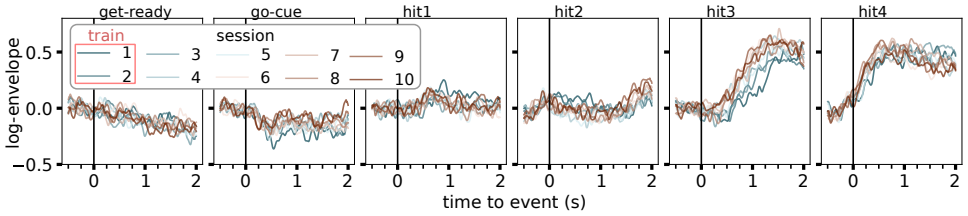


Figure 6.8: **Session-wise envelope dynamics of selected component along the sessions.** In analogy to Fig. 5.5, the session-wise averaged envelope dynamics aligned to various within-trial SVIPT events is exemplarily shown for the utilized beta-band component of patient *P*2 across all performed sessions. The corresponding filter model was trained on the first two offline sessions.

envelope of each session was baselined to a 500 ms interval prior to each event.

Remarkably, the envelope dynamics for this beta-band component is highly homogeneous across all (unseen) online sessions and shows distinct ERD/ERS modulations. As each online session provides novel data which was not used for the calibration, then assessing the envelope dynamics can be seen as a post-hoc generalization test of the *selected* subspace component. The stable ERD/ERS effects strongly advocate the functional relevance of the *selected* component. The ERD effect with *go-cue* and the ERS with *hit 3* suggests that the activation and deactivation of the sensorimotor system [240] along the trial is depicted with its full dynamics. The corresponding envelope dynamics of the remaining patients can be found in the appended Fig. A.6.

6.3.4 *Motor learning captured by clinical assessments*

In order to evaluate the individual clinical outcome of the brain state-dependent visuo-motor training, Tab. 6.3 show the three different clinical scores before and after training as well as the corresponding relative change.

Table 6.3: **Overview on clinical outcome metrics.** For all pilot patients, different clinical outcome metrics are reported for pre- and post-training assessments. Pinch and grip refer to measurements with their affected hand. Δ refers to the score-specific corresponding relative difference, taking the pre-training assessment as baseline.

	Pinch (N)			Grip (N)			UEFM (points)		
patient	*Pre*	*Post*	Δ (%)	*Pre*	*Post*	Δ (%)	*Pre*	*Post*	Δ (%)
*P*1	17.8	14.8	-16.7	35.6	29.7	-16.7	27.0	26.0	-3.7
*P*2	46.0	50.4	9.6	94.9	102.4	7.8	52.0	59.0	13.4
*P*3	23.7	28.2	18.8	20.8	60.8	192.9	53.0	58.0	9.4
*P*4	74.2	80.1	8.0	415.3	299.6	-27.9	58.0	60.0	3.4

Even though a statistical group-level evaluation is not feasible with the limited amount of pilot patients, one can still compare the relative changes

for all three scores on the level of individual patients to get a first impression. Interestingly, the patients *P*2 and *P*3 do reveal an increase in their pinch and grip force capabilities with their affected hand. Simultaneously, the patients show a gain in the UEFM score by at least 5 points which was reported as a clinically relevant improvement for mild to moderate motor impairments [241]. Likewise, the clinical outcome of *P*1 reports that the impairment exceeded a moderate level and suggests that SVIPT training might not be the most beneficial paradigm for such patients. Conversely, *P*4 was a non-naïve SVIPT user which minimized the overall motor learning. The individual clinical scores are also in good accordance with the SVIPT learning curve over the course of the 10 training sessions which is shown in the appended Fig. A.5.

6.4 DISCUSSION

In the following, the discussion of this chapter has a special role. Beside the detailed examination of the current chapter's findings, it also creates a link to all previous contributions in Chaps. 3, 4 and 5.

In summary, this chapter presented a novel brain state-dependent closed-loop interaction protocol exemplified for a repetitive hand motor training with chronic stroke patients. It is the follow-up online validation of the identified predictive oscillatory features that explain upcoming single-trial motor performance (see Chap. 3). In all four pilot patients, the developed methods were key to identify brain states that explain upcoming single-trial motor performance. Prior to the start of a single trial, a real-time brain state evaluation enabled to specifically select between *suitable* and *unsuitable* starting time points to direct the patients' upcoming motor performance. Based on individually optimized, task-specific and robust oscillatory features, reaction times under *suitable* brain states were on average shortened by up to 40 % of individually measured reaction time variability.

6.4.1 *Manipulating single-trial motor performance by the ongoing brain state*

The introduced closed-loop system to influence upcoming motor performance strongly depends on the evaluation of the ongoing brain state. Our system operated with an update rate of 40 *ms* which already resulted in a strong separation of oscillatory power between conditions.

Despite the challenging conditions with patients and comparably low predictive performances in the offline sessions (z-AUC between 0.62 and 0.69 as shown in Tab. 6.2), we found evidence to manipulate single-trial reaction times in all four chronic stroke patients by specifically selecting *suitable* and less frequent *unsuitable* starting time points for the online EEG-gated SVIPT motor task. Although we did not apply a brain state-dependent gating strategy in the offline sessions, we aimed for comparable conditions between the offline and online phase. Therefore, in each offline session the vertical cursor position was varied based on observed brain

state fluctuations of earlier pilot online sessions. As an additional control condition, we aimed to answer if there is already a split in reaction times during the offline sessions between the lowest and highest vertical cursor positions at *go-cue*. These positions referred to *suitable* and *unsuitable* trials in the later online sessions. Based on data of patients $P3$ and $P4$, where we extracted the corresponding markers, we did not find a significant separation of RT distributions in these cases. While this observation is a first indication that performance separation is not dependent on cursor position at *go-cue*, further data is required to fully conclude about this control condition.

For the majority of patients, we achieved a stable session-wise reaction time separation in the second half of the online training. This might be driven by the time to learn the reliable down-modulation of the pre-trial power. While we have not fully controlled for the volitionally induced modulation of the pre-trial power (versus a spontaneous fluctuation), an explicit SMR training over multiple sessions prior to the online phase was performed by Norman et al. [227]. In their pilot robot-assisted hand motor training with 8 chronic stroke patients [227], only 4 of them achieved the required reliable SMR control. In their online training over 3 sessions, finally in 3 out of 8 patients shorter reaction times were found when pre-trial oscillatory activity was low. Overall, the comparison of our study with [227] yields that we have performed more than the double amount of online motor training sessions—which supports a robust validation of the gating concept and enables the patient to have maximal motor training time—at the price of a reduced control over the pre-trial component power.

In accordance with earlier findings of three studies in total [226, 227, 242], it was found that a reduced SMR power correlated with shorter reaction times. Interestingly, this finding has now been shown in four different motor tasks (including SVIPT) with substantial variety in their underlying task complexity. While the three earlier studies could show the effect not even on half of their subjects, remarkably we have achieved the performance influence in all subjects. This finding advocates in favor of our data-driven framework for building the performance predictors. Although reaction time is not the final clinically relevant metric for post-stroke motor learning, it might provide a useful basic building block for the successful execution of higher-level sensorimotor tasks.

Even though the selected decoding model was optimized for reaction time, the training under *suitable* brain states also partially translated into enhanced motor performances captured by longer integration of behavioral performance. Interestingly, SVIPT trial durations were found to be significantly shorter for *suitable* than for *unsuitable* trials in two of four patients. Intuitively, the more behavioral information is integrated along the single trial, the less influence can be traced back by the brain state at *go-cue*. This finding is also in accordance with studies [226, 227] in which a behavioral performance split on metrics with longer integration intervals was only observed in a smaller fraction subjects. However, in our pilot data, we observed a separation on trial duration for the two patients ($P2$ and $P3$)

which showed the largest SVIPT motor learning effects along the training (see Fig. A.5) as well as the strongest clinical improvements (see Tab. 6.3).

6.4.2 *Calibration of the prediction model*

The calibration of a robust decoding model to predict upcoming single-trial motor performance is challenged by the amount of available training data which was limited to 200 trials per session. Due to the use of an online artifact detection pipeline in the pre-go phase, we successfully increased the number of available data points for model training after EEG preprocessing. However, the novel regularized NTik-SPoC method—introduced and evaluated in Chap. 4—requires about $N_{train} \geq 200$ to ensure a saturated decoding performance (see Sec. 4.4.1.3). Thus, two offline sessions were performed for the initial calibration with a chronological train/test-split to select an individual oscillatory component. A retraining of the model on all available labeled data was not performed to maintain feature introspection and not cope with rank instabilities (see Sec. 3.4.4). For two patients, $P3$ and $P4$, an initial training of the spatial filter models on all available channels resulted in a large fraction of artifactual components. As a mitigation strategy in such cases, we removed frontal channels for the training of NTik-SPoC and observed an increase in neural components.

Even though we utilized single predictive and functionally relevant components, from a machine learning perspective it could even be beneficial to combine regularized SPoC features across multiple frequency bands—comparable to the filterbank CSP approach [243]. This fusion of features can be achieved, e.g., by an additional regression model. This might allow for enhancing the trial-wise performance prediction, given that the information contained in different frequency bands is independent. Similarly, the combination of predictors based on different performance metrics might serve to gain an enhanced predictive power.

6.4.3 *Careful adaptation of the prediction model*

As non-stationarity effects in recorded brain activity generally impede a robust decoding [238], an online adaptation was applied. Even though each online session continuously provided novel labeled data and thus a constant re-calibration after each session could have been performed, here only a careful model adaptation was carried out. In fact, the selected spatial filter model was fixed after the initial calibration. By the continuous but refined adaptation of the gating thresholds, we could successfully achieve the targeted gating ratio across the full training for each individual patient. The control over this ratio is of specific importance if the influence of different gating strategies on post-stroke motor learning is studied. As previously argued by Biasiucci et al. [16], a repeated re-calibration during the BCI training, such as realized by Ang et al. [244], might translate into different oscillatory components with diverging efferent pathways and thus hinder training-induced plastic changes. Overall, the degree of model

adaptation over the course of a closed-loop training is an ongoing debate and requires further investigation.

6.4.4 *Spatial filters allow for in-depth introspection*

The trained data-driven spatial filters for the performance prediction allow for a comprehensive analysis of selected oscillatory features. As provided in Sec. 6.3.3, across-session spectral and topographical characteristics can directly be assessed by the deployed spatial filter models. The spectral and topographical analysis contributes towards the neurophysiological understanding of the targeted oscillatory features. In this pilot study, we identified and utilized predictive features in alpha- and beta-band frequencies which were reported to be crucial in the preparation and execution of accurate motor performance [245, 246].

As a central novelty, the within-trial SVIPT envelope dynamics were intensively exploited for the performed online experiments. The introduced framework in Chap. 5 contributed to this application in two ways: first, the mining approach on within-trial SVIPT envelope dynamics facilitated the selection process of individual robust and functionally relevant oscillatory components for the closed-loop interaction. However, keeping a neurophysiological expert for the final feature selection in the loop is (still) a common strategy for BCI-based trainings [16, 74, 247]. Beside the feature's predictive power, we took a variety of feature characteristics into account, such as spectral and topographical information as well as the envelope dynamics. Such an in-depth feature introspection path finally grounds the component selection on diversified sources of information.

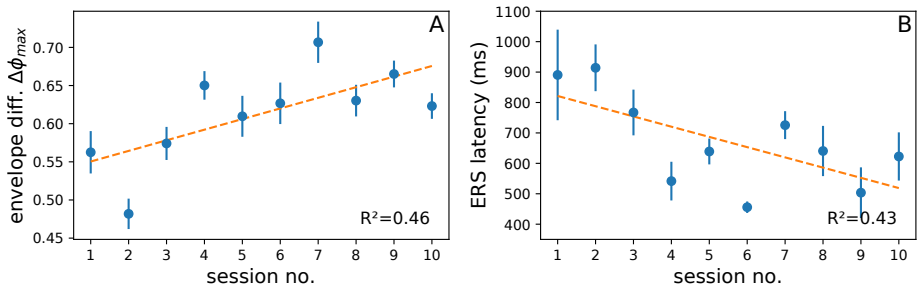


Figure 6.9: **Component-specific changes in ERD/ERS characteristics across the training sessions for patient** *P2***.** As a follow-up analysis based on results in Fig. 6.8, the component-specific envelope difference (A) and the corresponding ERS latency (B) after *hit 3* is plotted over the course of the training. For each reported data point, the mean value and corresponding standard deviation resulting from a bootstrapping procedure is reported. Both plots are complemented by a linear fit (dashed lines) and the corresponding R^2 value.

Second, the concept of taking the within-trial envelope dynamics as a central source of information enabled for a direct evaluation of expected training-induced functional changes. Precisely, changes in the temporal envelope dynamics of patient-specific functionally relevant components

over the course of the hand motor training become accessible, as reported in Figs. 6.8 and A.3. As such, differences in amplitudes or timings of ERD/ERS effects can be quantified as exemplarily shown for the selected oscillatory component of patient *P*2 in Fig. 6.9. Therefore, event-specific ERD/ERS envelope difference and latency was estimated in a bootstrapping procedure with 20 repetitions. For each bootstrap sample, the session-average event-related envelope dynamics was calculated on 75 % of the session-specific data. Then, the event-specific maximum envelope difference (as introduced in Sec. 5.2.3.5) as well as the corresponding *latency* was determined. The latency was defined by the time interval from the event onset to the time point in which 4 times the baseline envelope variation (measured in an interval 500 ms prior to the event) was de- or exceeded.

For the beta-band component of *P*2, in all single SVIPT sessions a consistent ERS effect can be observed after *hit 3* due to the positive maximal envelope difference $\Delta\phi_{max}$. Interestingly, over the course of the training there is a significant increase in the ERS amplitude as well as a significant reduction of the corresponding ERS latency. Such training-induced changes of sensorimotor bandpower features across sessions were reported earlier in ERD-BCI paradigms, as users gain increased control over the BCI [248] or modulated oscillatory features in a MI-based stroke rehabilitation training [74].

In summary, the proposed methodology shown in Fig. 6.9 can be applied and extended to any SVIPT event and functionally relevant component. This might pave the way for novel assessment tools to track and monitor expected training-induced functional changes in individual patients.

6.4.5 *Training under desired brain states might enhance post-stroke motor learning*

Exemplified for SVIPT hand motor training, we provided a temporal gating strategy on a millisecond time scale which can be further used to adapt the task difficulty to the current brain state. In this pilot study, the patients trained mostly under *suitable* brain states and we solely reported on cases in which a low oscillatory power at *go-cue* was associated with a high motor performance, particularly shortened reaction times. Brain states of low oscillatory power prior to a movement might foster the cortex excitability as well as the ability to generate and process afferent input [249].

For two patients we even found that *suitable* pre-trial brain states translated into significantly reduced SVIPT trial durations. Thus, a subsequent research question would be to investigate if repetitive motor training under desired brain states is then also beneficial for motor learning after stroke. As done in the pilot study, one possible approach would be a facilitation strategy by providing mostly *suitable* brain states. Due to the small sample size, the effect on motor learning cannot be answered so far.

However, comparable to bilateral priming, a technique which was shown to enhance the excitability of the motor system and subsequently accelerate motor learning after stroke [250], a similar effect could be expected for the

gating concept such that it could elicit desirable motor preparation states on a shorter millisecond timescale.

6.4.6 *Applicability of brain state-dependent gating*

The provided brain state-dependent gating concept is complementary to most existing BCI-based systems in the field of post-stroke motor rehabilitation [15, 74]. These systems focus on the direct decoding of movement intentions, while the gating concept is taking trial-wise brain state fluctuations into account which influence upcoming motor performance. Thus, the gating concept could prove beneficial for other repetitive motor paradigms in post-stroke rehabilitation.

Conceptually, the work provided in this chapter can be seen as a sample application for brain state-dependent experimenting [20]. As our framework is based on a data-driven decoding model without prior assumptions on the underlying cortical network, it is not limited to a specific application. It only requires single-trial labels of behavioral variability to identify a corresponding neural correlate. As previously reported in Sec. 3.4.5 and 5.3.1.1, our framework might allow to assess different cognitive sub-processes, which are related to, e.g., an ongoing attentional level. Thus, there are various application fields beyond motor rehabilitation. As an example, we foresee that the brain state-dependent gating concept could be beneficial in cognitive rehabilitation scenarios such as in attention-related disorders as attention deficit hyperactivity disorder (ADHD) [251]. Moreover, sports science could profit from the consideration of ongoing brain states, specifically for the development of training concepts in which single-trial performance should be optimized. These aspects might play an important role in disciplines such as archery, darts or ski jumping.

6.4.7 *Outlook*

After successfully demonstrating the feasibility to manipulate upcoming single-trial motor performance based on the contributed building blocks of this thesis, a variety of further interesting research questions come up.

From a clinical perspective, one can now verify if such a brain state-dependent post-stroke motor training is more effective for motor skill learning than established training protocols without any selection of desired brain states [17]. In this respect, I contributed substantially to a successful grant proposal to test the brain state-dependent training efficacy and efficiency in a randomized controlled trial on a larger sample size.

For the patient training, a major challenge is to decide about the potentially most beneficial gating strategy for the brain state-dependent motor training. So far, only little is known about the influence of task difficulty on post-stroke motor learning [102]. Thus, we do not know a priori to which extent a defined gating ratio affects post-stroke motor learning. For each specific gating strategy, one would need to run a full study to evaluate the underlying efficacy. Based on the current findings, the author suggests run-

ning a training under mostly *unsuitable* brain states to carefully challenge the patient in most trials of the training [252]. Still, a fraction of *suitable* trials would be favorable both for the motivational level of the patient and to continuously gain labels of both conditions.

Going beyond the proposed *passive* gating strategy, a further development would be to study the *active* triggering of desired brain states, e.g., by means of functional electrical stimulation [16].

From a computer science point of view, this specific clinical application bears a number of interesting challenges:

The efficacy of the closed-loop gating strategy strongly depends on the accuracy of the individually trained prediction model. To further improve the finding of robust and functionally relevant features and eventually save calibration time, various across-subject transfer learning strategies could prove beneficial as they revealed promising results in comparable BCI scenarios [62, 146]. An improvement in decoding accuracy could also be achieved by involving further task-relevant features beyond oscillatory power. Among others, phase-related information could be of particular interest [44, 253]. Likewise, the consideration of deep learning strategies might provide added value for performance prediction [1, 100], especially if larger data collections are available. Finally, the real-time power estimation could be further optimized by identifying optimal window sizes for power estimation or by deploying stochastic methods for selecting optimal brain states [254].

Regarding the feature selection strategy for the closed-loop interaction, a fully automatic way to select the most promising functionally relevant and predictive feature can be established in the future. For such a system, it might be beneficial to define a set of metrics which express all feature requirements and thus allow for an objective choice.

In the reported multi-session scenario, non-stationarities of the selected oscillatory components could be characterized in detail for the design of novel online adaptation strategies that directly cope with the observed fluctuation characteristics. Also, the timing and strength of online adaptation for closed-loop experimentation is only partially studied so far and requires further investigation [16].

6.5 LESSONS LEARNED

This chapter fuses all previous contributions of Chaps. 3, 4 and 5 into a final closed-loop application: in a nutshell, the feasibility of a novel online temporal gating strategy to influence upcoming single-trial motor performance was successfully established. This performance influence was demonstrated in a pilot multi-session hand motor training with four chronic stroke patients which revealed strong trial-to-trial motor performance variations. Even under challenging conditions with patients, we could identify robust and predictive brain states that allowed the gating of *suitable* pre-trial starting time points of an upcoming motor task. Those elicited an increased upcoming motor performance. Particularly, single-trial

reaction times were significantly reduced—ranging from $27 - 40\,\%$ of the individual reaction time variations—for *suitable* trials than for *unsuitable* labeled starting time points. As this framework on real-time brain state interaction is not exclusively designed for motor rehabilitation, the detection and exploitation of *un-/suitable* brain states can potentially be transferred to different applications such as cognitive trainings or sports sciences.

Part III

FINAL SUMMARY OF THE CONTRIBUTIONS

7 SUMMARY AND OUTLOOK

This section summarizes the overall contributions of this thesis and relates them to the originally posed research questions (see Sec. 1.1). As sketched in Fig. 7.1, the major findings of the previous chapters can be wrapped up into two main categories: (1) methodological contributions for robust brain state decoding which are of general relevance to the field of BCI research and (2) application-related findings in the context of the newly introduced brain state-dependent interaction strategy exemplified for a post-stroke hand motor training. Finally, a brief outlook is provided focusing on the possible impact on future research.

7.1 SUMMARY OF CONTRIBUTIONS

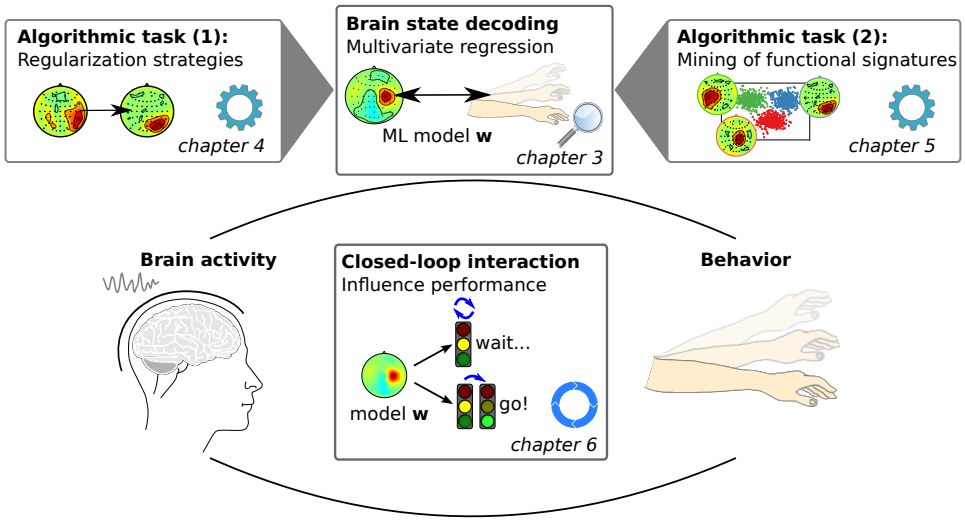


Figure 7.1: **Conceptional overview on thesis contributions.** The single contributions of the thesis are schematically visualized and related to the final closed-loop application scenario.

Two algorithmic contributions for single-trial brain state decoding

From a machine learning point of view, this dissertation pushed forward two types of algorithmic contributions that foster robust single-trial brain state decoding under challenging data regimes, as typically encountered for many BCI applications [18, 62]. As this thesis additionally focused on a final clinical application scenario, strong emphasis was also put on a model's ability to allow for detailed feature introspection.

The first algorithmic contribution (see Chap. 4) comprises regularization techniques for the spatial filtering algorithm SPoC [94]. This supervised multivariate approach allows regressing a continuous variable from oscillatory multichannel EEG activity. However, in the above described data regimes the non-regularized SPoC method is prone to overfit to the training data. To improve SPoC upon this challenge, different types of regularization techniques were proposed and evaluated both in a novel simulation framework and on real-world datasets, respectively. Confirming our expectations, SPoC regularization generally reveals the largest benefit under small training datasets and severe label noise conditions [32, 33]. Overall, a variant including a cross-validation based Tikhonov regularization and additional covariance normalization turned out to be the most beneficial technique. The introduced framework for the evaluation and benchmarking of various regularization strategies allows validating further, e.g., application-driven regularization strategies. Finally, regularized SPoC variants that are robust in small data scenarios are of particular interest in various closed-loop applications where *continuous* brain state estimates on single trial level are necessary. Examples are the estimation of ongoing workload levels in real-world scenarios [179, 180], the depth of cognitive processes [181] or movement-related information [182].

The second algorithmic contribution (see Chap. 5) addresses the variability of optimized spatial filters, a widely used class of machine learning models in BCI applications [53]. However, in most real-world data scenarios trained spatial filter models are extremely sensitive to slight changes in training data or involved hyperparameters, such as frequency bands, time intervals or regularization parameters [31]. This sensitivity leads to highly variable filter solutions and impedes the selection of a suitable candidate for, e.g., neurotechnological applications. Moreover, traditional BCI applications are typically tuned to solely maximize the *decoding accuracy*, while disregarding model interpretability or even the functional role of oscillatory features [90]. This gap was closed by exploiting the within-trial temporal dynamics of single oscillatory components. Precisely, I introduced a novel method that embraces the observed filter variability by condensing the functional signatures of a large set of oscillatory components' into homogeneous clusters, each representing subject-specific within-trial envelope dynamics [34]. The method was evaluated on data of two different motor paradigms with a rich within-trial structure but of diverging task complexity. For both scenarios, I found that the components' distinct temporal envelope dynamics are highly subject-specific and strictly confined regarding their underlying frequency band. As the analysis method is not restricted to a specific spatial filtering algorithm, it could potentially be utilized to select *reliable and functionally relevant* features for a wide range of neurotechnological applications. A prominent application example is to support the process of selecting and monitoring features for BCI-based protocols in stroke rehabilitation [10, 15, 74].

In a nutshell, the two algorithmic contributions overall improved on the robustness and reliability of existing brain state decoding algorithms,

as originally motivated by my research questions Q2a and Q2b. Both algorithmic building blocks share a more general applicability for various neurotechnological systems.

Oscillatory features enable for single-trial performance prediction

The thesis is complemented by two additional contributions that focus on a real-world application scenario: I studied a repetitive hand force task, named SVIPT [115], which was originally designed for post-stroke motor learning. As a novelty, the paradigm was equipped with an additional tracking of brain activity with a 63 channel EEG system. Overall, datasets of two studies—one with chronic stroke patients and the other one with normally aged control subjects—were recorded for offline analysis.

A behavioral data analysis revealed strong trial-by-trial motor performance variations within single sessions, especially for the motor impaired patients. As captured in my research question Q1, it was investigated whether individual oscillatory brain activity [27] can provide valuable information about the upcoming behavioral performance (see Chap. 3). To tackle this question, the above mentioned regularized SPoC algorithm was deployed to regress various *continuous* motor performance metrics, such as reaction times, path lengths or jerk related metrics, from multichannel oscillatory brain activity recordings.

In an offline analysis, subject-specific single predictors that explain up to 36 % of the observed performance fluctuations were identified [28, 29]. From a machine learning perspective, a set of robustness criteria was developed that allows identifying robust predictive oscillatory features. Such meaningful predictors were found both on data of normally aged controls and on data of chronic stroke patients [142]. They might reflect cognitive sub-processes that are relevant for successful preparation to the visuo-motor task. Overall, reaction time turned out to reveal the largest decoding performances among the investigated performance metrics.

Brain state-induced motor performance manipulation

The offline findings on predictive oscillatory features were finally transferred into a closed-loop experimental protocol as sketched in Fig. 7.1. Given the two developed algorithmic advances at hand, they were indispensable building blocks to finally establish and validate a novel brain state-dependent interaction strategy (see Chap. 6):

As large behavioral trial-to-trial performance fluctuations occur in scenarios like a post-stroke motor training, I verified if the real-time estimation of a predictive pre-trial brain state can be utilized to influence upcoming motor performance on single-trial level, as stated in my original research question Q3. The performance influence was accomplished by a temporal gating strategy prior to the start of a trial. A SVIPT trial start was mostly elicited if a user-specific performance prediction model indicated a desired brain state. In a proof-of-concept study, I demonstrated the feasibility of

this concept in a high intensity hand motor training on four chronic stroke patients—with $\approx$ 15 hours of effective training per patient. Even under challenging conditions with patients, I could identify robust brain states that allowed the gating of *suitable* pre-trial starting time points, which elicited an increased upcoming motor performance compared to *unsuitable* starting times. Noticeably, in all four patients single-trial reaction times were significantly reduced—ranging from 27–40 % of the individual reaction time variations—for *suitable* trials compared to *unsuitable* labeled starting time points.

This gating approach can be categorized into the recently introduced research branch of brain state-dependent experimenting [20]. Complementary to most BCI systems in which a direct decoding of user intention is realized [12], here the focus was put on the active exploitation of *additional information* about a user's current brain state—as thoroughly discussed in Sec. 6.4.

7.2 OUTLOOK

Finally, I want to briefly highlight the most promising future directions that directly build upon the contributions of this thesis. A more in-depth outlook can be found in Sec. 6.4.7.

Clinical applications and beyond

From a clinical point of view, it will be important to investigate if the introduced brain state-dependent gating strategy can be effectively deployed for motor training scenarios. Taking post-stroke rehabilitation as an application field, it is mandatory to evaluate whether a brain state-dependent post-stroke motor training is more effective for motor learning than the established training protocol without temporal gating. To study this question adequately, I contributed to a successful grant proposal to finally test the brain state-dependent training efficacy and efficiency in a randomized controlled trial on a larger sample size.

Likewise, the brain state-dependent adaptation of difficulty levels for forthcoming tasks enable for novel application scenarios as the introduced decoding framework was designed to come up with any informative neural sub-process—not at all restricted to a certain cortical area—that holds relevant information about performance variations. Possible example application fields might be: (1) cognitive rehabilitation trainings, e.g., for people with attentional deficits as in ADHD and (2) novel training concepts in sports sciences in which specific aspects of motor performance are optimized as highly relevant for disciplines like archery or ski jumping.

Algorithmic challenges

The above stated application fields yield novel challenging algorithmic questions that could be tackled in future research:

First, the efficacy of brain state-dependent training systems are strongly influenced by finding informative, robust and functionally relevant features. To accomplish this challenge with minimal calibration effort [62], the decoding methods for performance prediction could potentially be improved by applying transfer learning strategies which accumulate information across sessions, subjects or even experimental tasks. At the same time, this procedure would additionally promote the transfer of BCI systems from the lab to home-use environments.

Second, the decoding of functionally relevant features or sub-processes could be deepened. Especially in complex tasks, the interaction of various neural sub-processes are necessary building blocks for evoking a favorable behavioral performance. Following the proposed path of exploiting within-trial brain signatures, the tracking and monitoring of observable sub-processes during motor tasks as well as a statistical evaluation of timings and interactions could be a valuable future track. In a next step, it might be beneficial to utilize non-linear decoding approaches, such as deep learning methods [1], which could aim to disentangle functionally relevant sub-processes and their interactions from brain activity recordings. This could be accomplished by directly integrating more fine-granular information from complex tasks into decoding methods. This course of action might contribute to the general understanding of the motor system as well as motor learning, when comparing the findings of the novel approaches on data of healthy controls and patients. By that, one might gain access to, e.g., continuously monitor the patient's rehabilitation progress based on non-invasive brain activity recordings.

FINAL STATEMENT

To conclude, my dissertation sets an example for interdisciplinary work at the junction of computer science, clinical research and basic neuroscience. The achieved algorithmic developments finally made it possible to establish and conduct a novel neurotechnological application—a brain state-dependent post-stroke motor training protocol. Interestingly, having such a concrete application scenario at hand, a variety of machine learning challenges come to light. Moreover, the present thesis has shown that improvements in a final (clinical) application were enabled by a holistic view across different disciplines. In the future, more cross-domain efforts of this kind could also be promising for other application scenarios.

Part IV

APPENDIX

APPENDIX

A.1 SUPPLEMENTARY MATERIAL TO CHAPTER 3

Single-trial envelopes of exemplary components

The computed spatial filters can be applied on single-trial level to verify their predictive strength. Fig. 3.6 reveals the single-trial envelopes of two exemplary oscillatory components. In order to emphasize the oscillatory components' ability to be informative about upcoming motor performance, the single-trial envelopes are divided in trials of the upper and lower RT distribution. For both examples it can be nicely seen that a large pre-go power is dominated by the "high RT" class and vice versa. Interestingly, in (B) the role of the timing of the components' rhythmic activity can be well observed. While envelopes of the "low RT" class reveal a high power in an interval around $[-2, -1]\,s$ and then decays to lower powers in the gray shaded interval, the "high RT" class reveals the largest power in the interval used for training. This supports our finding, that the actual pre-go bandpower of an oscillatory components partially explains upcoming single-trial motor performance.

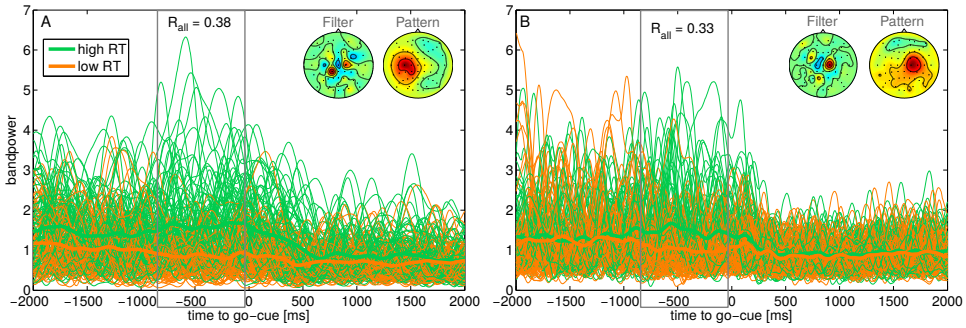


Figure A.1: **Single-trial envelopes of predictive components grouped into extreme single trial motor performance.** The exemplary SPoC components in (A) and (B) were trained on EEG data of the interval $[-800, -50]\,ms$ prior to the go-cue (gray boxes). Each line refers to a single-trial envelope of the corresponding component. In green, all envelopes corresponding to trials in the upper RT quartile are shown, while in orange the trials related to lower quartile RTs are reported. The thick lines correspond to the averaged envelope for the corresponding quartile. The component shown in (A) is derived from subject S13 in the alpha-band range, while (B) is the same configuration of subject S5 which was reported in Fig. 3.6.

Single-trial predictors of chronic stroke patients

In Fig. A.2, two exemplary predictive oscillatory SPoC components gained on the dataset (D1b) of chronic stroke patients are shown. The example of *P*2 was gained from the alpha band, the one of *P*6 was found in the beta-band.

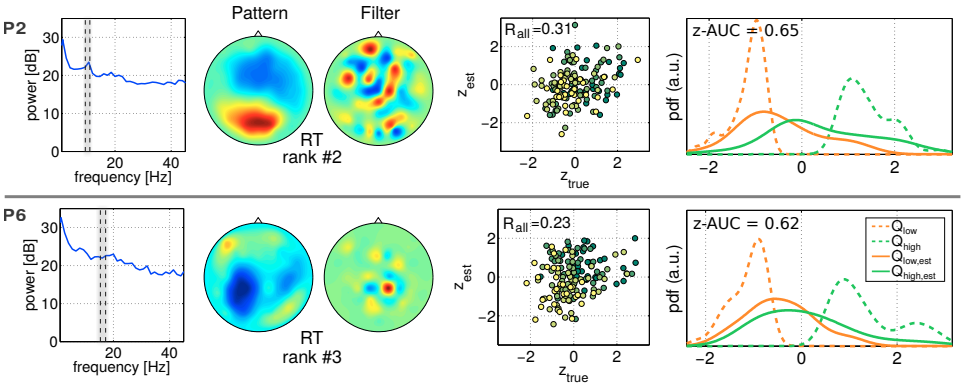


Figure A.2: **Exemplary predictive SPoC features of chronic stroke patients.** The presented components were gained on the single sessions of dataset (D1b). Each component is characterized line-wise labeled by the used performance metric and the rank according to the full-session filters. The same characteristics as in Fig. 3.6 are reported.

A.2 SUPPLEMENTARY MATERIAL TO CHAPTER 4

Effect of regularization upon oscillatory components

For an exemplary subject, the effect of the regularization strength in NTik-SPoC upon the underlying first four ranked oscillatory components is depicted in Fig. A.3. In (A), the z-AUC is reported for the different α values, in (B) the first four ranked patterns of the marked evaluation point (a)-(i) are shown. As the sign of a pattern $\mathbf{a}$ is arbitrary, they have been corrected to be consistent across the displayed patterns and scaled by their norm.

Consistent with the results presented in Sec. 4.4.1.3, three different α ranges can be identified for the selected subject. They can be characterized according to two aspects: the performance (A) and the underlying spatial patterns (B).

For very small regularization values, represented by evaluation points (a) and (b), the performance as well as the spatial patterns are stable despite of increased α. In other words, NTik-SPoC is not sensitive for such small values of α. The gray shaded area in Fig. A.3 (A) encloses the evaluation points (c)–(f). This range is sensitive to changes of α which is revealed by performance improvements as well as rank switches among the spatial patterns (e.g. rank #4 from (d) tracked by a solid red line to (f)) or even novel patterns that appear among the top four ranks (e.g. rank #3 at position (d)). In Fig. A.3 (B), novel patterns are marked by a red circumference.

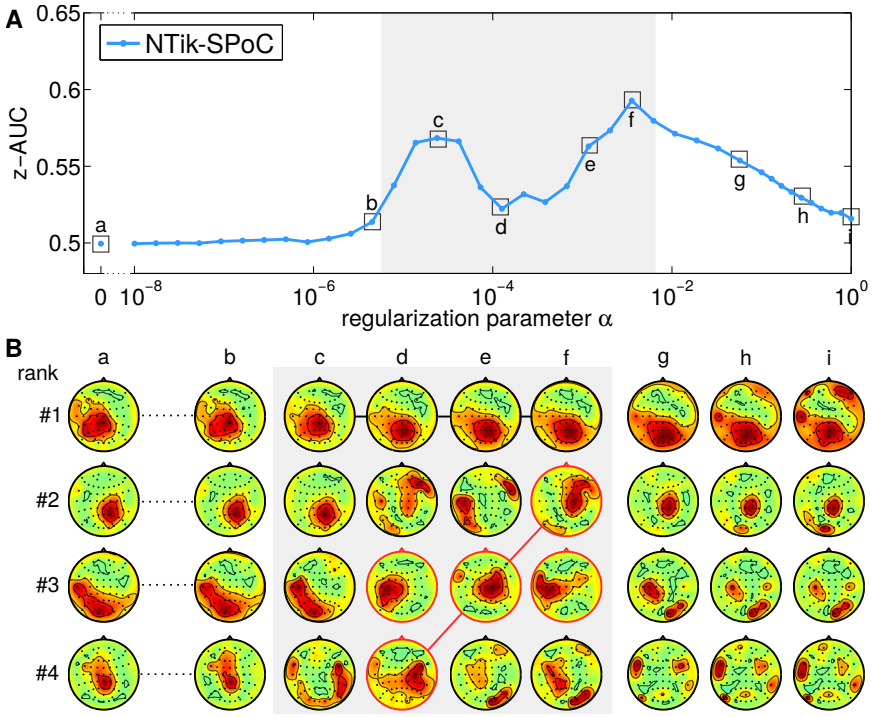


Figure A.3: **Example to visualize the effect of regularization strength on performance and the spatial patterns.** The shown results are reported for an exemplary subject S3. (A) Performance of NTR-SPoC for choosing $\alpha \in [10^{-8}, 10^{0}]$. (B) Corresponding activity patterns along first four ranked components at the marked evaluation points (a)-(i).

Regularization beyond $\alpha > 10^{-2}$ (as in (g)–(i)) leads to a slight drop in performance. This is accompanied by an increased number of components among the first ranks, which display higher spatial frequencies in their activation patterns. The latter observation was made for most subjects and usually affected patterns of ranks 2–4.

A.3 SUPPLEMENTARY MATERIAL TO CHAPTER 5

Envelope dynamics of clusters

To demonstrate the applicability of the mining framework also for the patient scenario with scarce training data, Fig. A.4 gives a detailed overview on representative clusters of oscillatory NTik-SPoC components that were gained on the SVIPT dataset (D1b) which contains data of chronic stroke patients. Comparable to Fig. 5.5, each row (C1)–(C7) contains an exemplary cluster and its set of associated within-trial event-related envelope traces. The examples were again selected to represent typically observed effects related to band-specific amplitude modulations, underlying frequency ranges and cluster homogeneity. Specifically, rows (C1)–(C3) refer to clusters of patient P3, while (C4)–(C5) are gained from patient P4 and (C6)–(C7) correspond to P7.

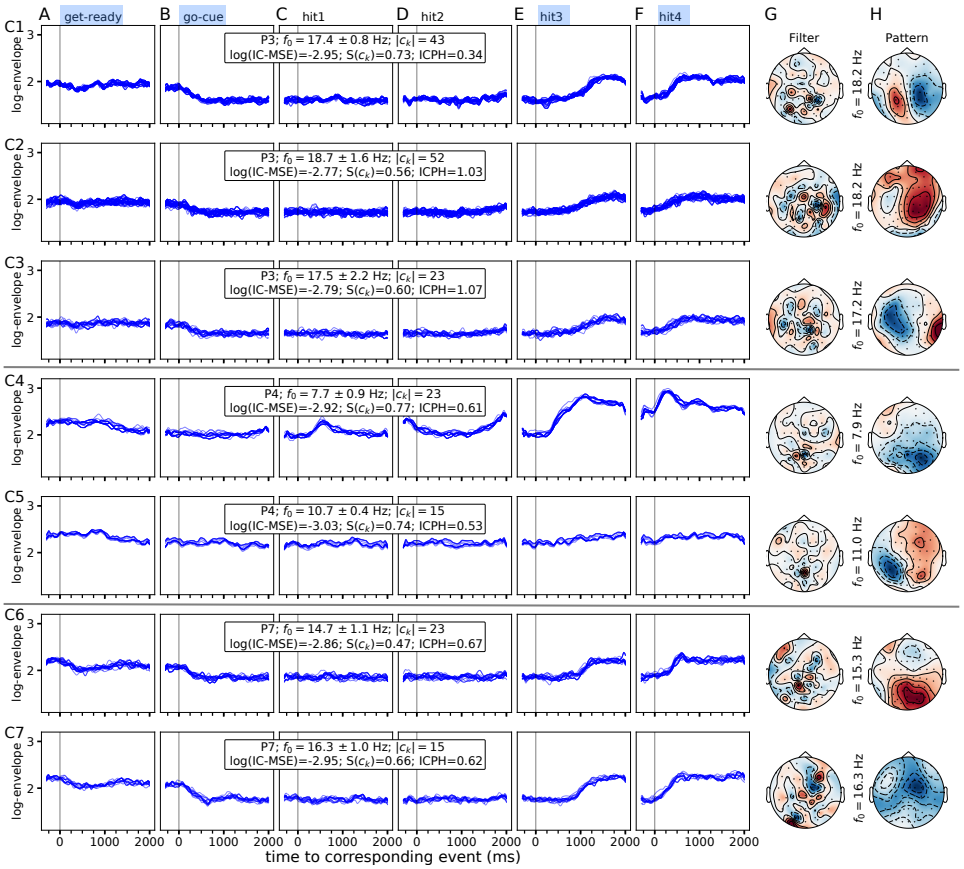


Figure A.4: **Representative envelope dynamics for individual clusters on data of chronic stroke patients.** All configurations for single subject-specific clusters are reported in rows (C1)–(C7). Columns (A)–(F) report the cluster-wise envelope dynamics for within-trial SVIPT events, while in (G) the spatial filter and in (H) the related activity pattern of cluster representatives (with annotated central frequency) are shown. In all subplots of columns (A)–(F), every blue line refers to the log-envelope dynamics $\phi_j(t, m)$ of one single hyperparameter configuration $\omega_j \in c_k$. Only events shaded in blue were included for the clustering step. The text box on top of each row provides the subject code, the mean and standard deviation of the central frequency across all cluster samples, the cluster size, the average decoding performance as well as three validation metrics.

In accordance with the findings based on data of normally aged subjects, the envelope dynamics aligned to the different within-trial SVIPT events reveal distinct and time-locked ERD or ERS effects. Again, ERD effects can be found for *get-ready* and *go-cue*, while *hit3* and *hit4* elicit ERS effects.

The analysis also confirms that the event-related envelope dynamics reveal substantially different shapes both within and across subjects. To state an example, the two clusters (C1) and (C3) of the same patient are both associated with the beta-band domain, while their envelope dynamics shows distinct differences over time, such as the diverging ERS effects towards the end of the trial. Observable ERD/ERS effects are remarkably different across subjects and clusters.

This demonstrates that the proposed framework also allowed to identify individual clusters of homogeneous envelope dynamics in the patient scenario. This is an important prerequisite for the method's applicability in the preparation of closed-loop scenarios.

A.4 SUPPLEMENTARY MATERIAL TO CHAPTER 6

SVIPT motor learning

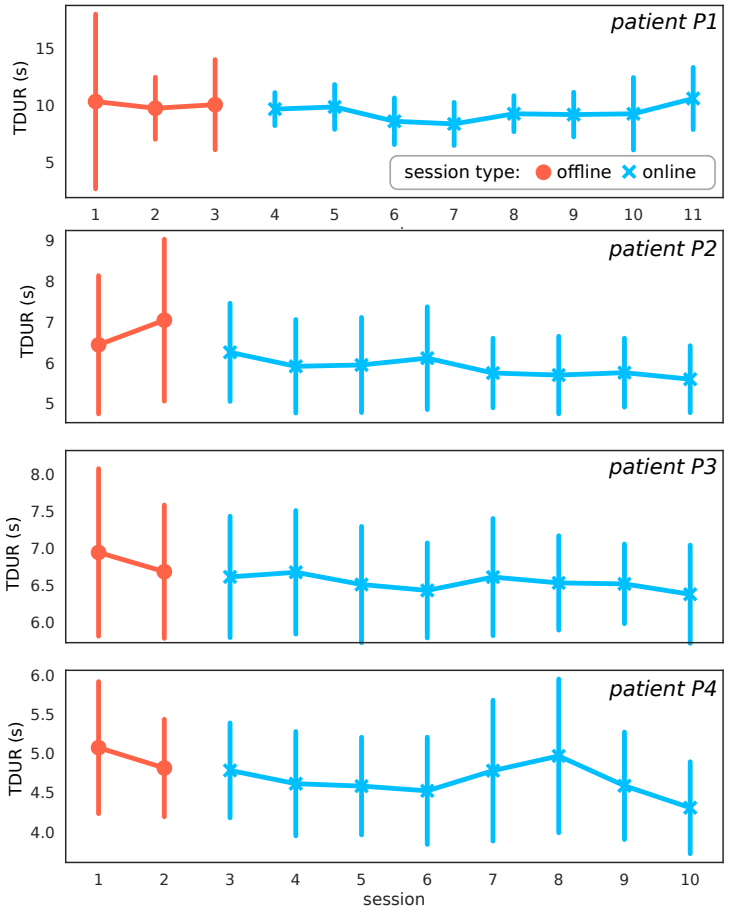


Figure A.5: **SVIPT motor learning over the course of the hand motor training.** For each pilot patient, the session-wise median of the SVIPT trial duration and the corresponding standard deviation is shown.

The patient-specific SVIPT motor learning over the course of the intensive training is quantified in Fig. A.5. For patients $P2$ and $P3$ a continuous decrease of the average SVIPT trial duration can be stated accompanied by a reduction of performance variation over the course of the training. The remaining patients reveal a different behavior. $P1$ reveals larger overall trial durations and a rather inconsistent behavior from session-to-session. This can be attributed to the severity of his motor impairment (see UEFM in Tab. 6.3). In contrast, $P4$ was a non-naïve SVIPT user and less motor im-

paired as all other patients. As expected, the course of SVIPT performance from session-to-session did not reveal a strong motor learning effect.

Oscillatory feature introspection

As the introduced decoding framework allows for a post-hoc inspection of the utilized oscillatory features for closed-loop interaction, a patient-specific summary of the underlying event-related envelope dynamics (for details see chapter 5) is given in Fig. A.6. Overall, it can be observed that the selected features reveal stable characteristics along the training, highlighting a differently pronounced ERD with *go-cue* and an ERS with *hit 3* (for *P*2) or *hit 4* (for all other patients).

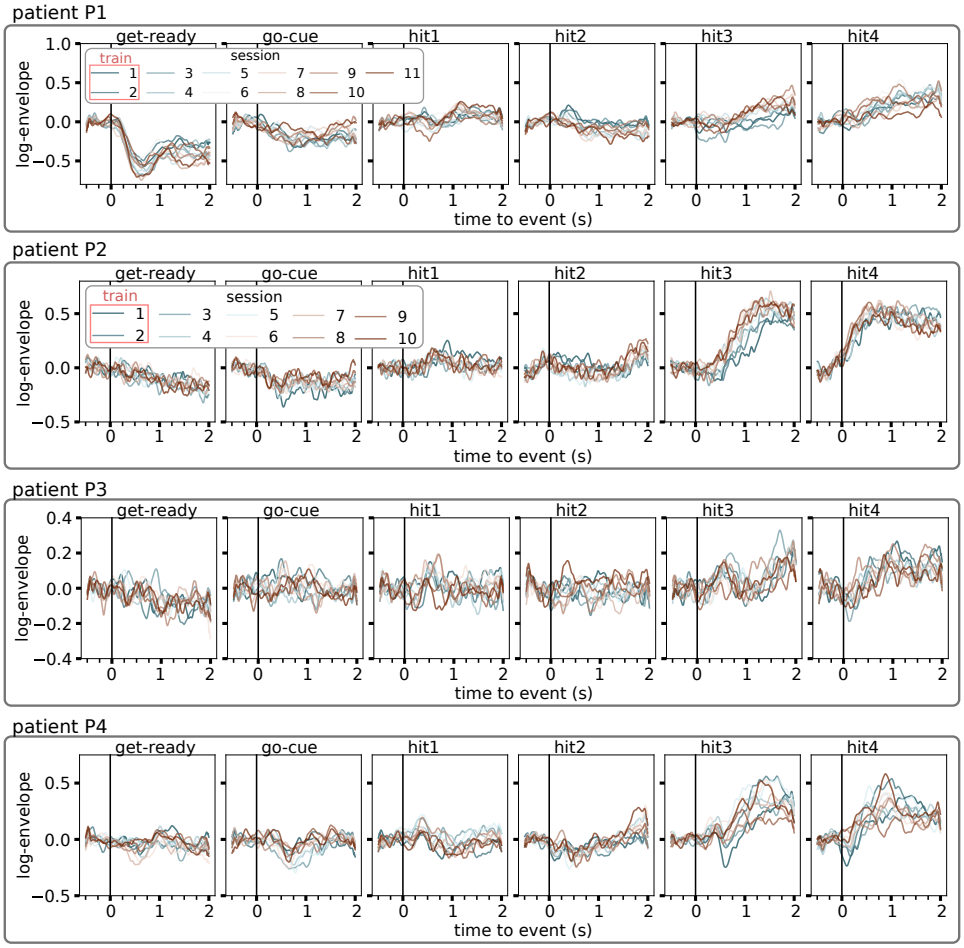


Figure A.6: **Session-wise within-trial envelope dynamics of individual features for closed-loop interaction.** For all pilot patients, the session-averaged envelope dynamics aligned to the different within-trial SVIPT events are shown for single training sessions. Patients *P*2-*P*4 share the same legend.

In addition, an overview on the session-wise pre-trial spatial activity patterns of the underlying spatial filter model are reported for all pilot patients in Fig. A.7. It can be observed that the patterns behave differently across patients. While for *P*3 and *P*4 a mostly stable spatial topography

from session-to-session is found, for $P1$ and $P2$ perceivable variations are contained across sessions. These variations might be due to signal non-stationarities, e.g. different impedances at electrodes from session-to-session or systematic changes in the underlying brain activity.

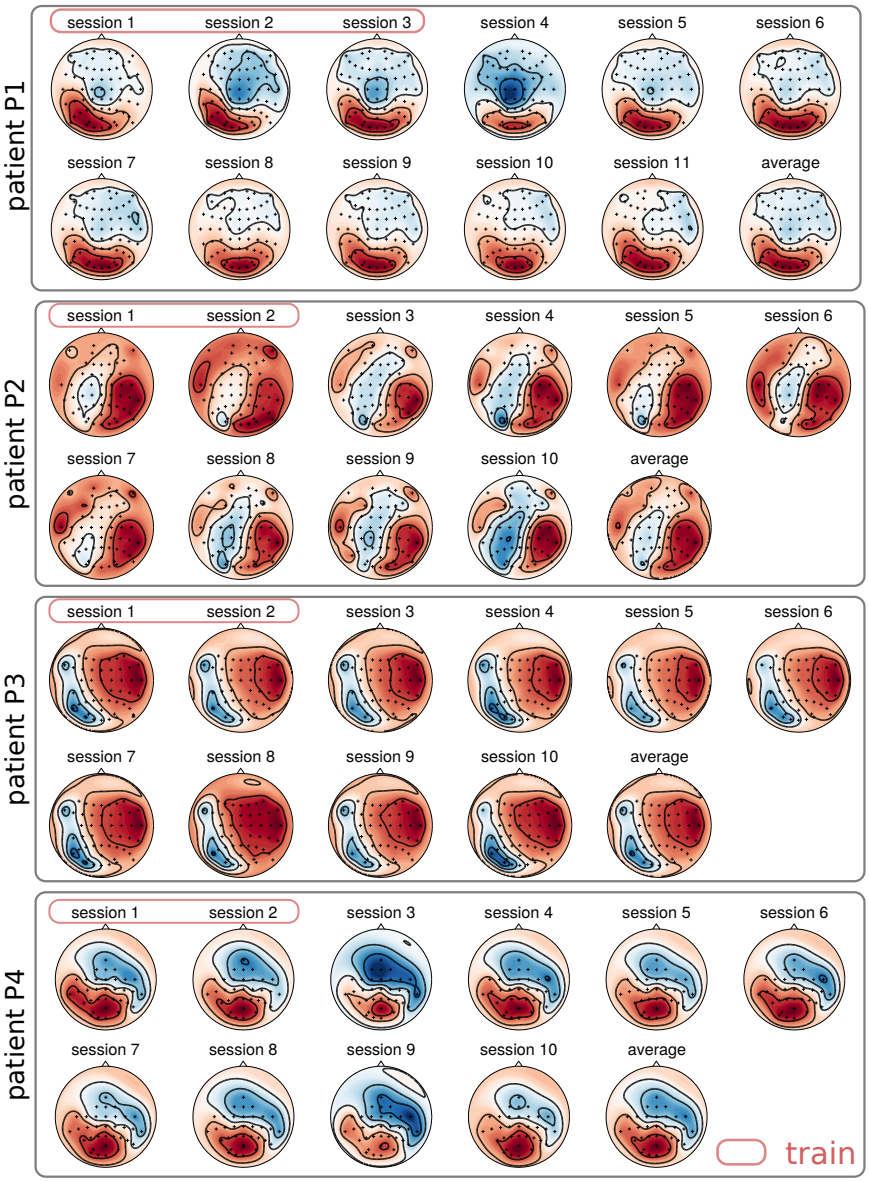


Figure A.7: **Pre-trial spatial activity patterns of selected features.** For all pilot patients, the pre-trial activity pattern of the underlying oscillatory component is reported for each session. In addition, the average pattern across all sessions is provided. For patients $P1$ and $P3$, the lastly selected oscillatory component is shown (according to Tab. 6.2). Please note that the corresponding filter was only trained on the first two (for $P1$: three) sessions.

BIBLIOGRAPHY

[1] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli. "Applications of Deep Learning and Reinforcement Learning to Biological Data". In: *IEEE Transactions on Neural Networks and Learning Systems* 29.6 (June 2018), pp. 2063–2079.

[2] M. W. Libbrecht and W. S. Noble. "Machine learning applications in genetics and genomics". en. In: *Nature Reviews Genetics* 16.6 (June 2015), pp. 321–332.

[3] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis. "Machine learning applications in cancer prognosis and prediction". In: *Computational and Structural Biotechnology Journal* 13 (Jan. 2015), pp. 8–17.

[4] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. "Dermatologist-level classification of skin cancer with deep neural networks". en. In: *Nature* 542.7639 (Feb. 2017), pp. 115–118.

[5] S. Kiranyaz, T. Ince, and M. Gabbouj. "Personalized Monitoring and Advance Warning System for Cardiac Arrhythmias". En. In: *Scientific Reports* 7.1 (Aug. 2017), p. 9270.

[6] K. R. Foster, R. Koprowski, and J. D. Skufca. "Machine learning, medical diagnosis, and biomedical engineering research - commentary". In: *BioMedical Engineering OnLine* 13.1 (July 2014), p. 94.

[7] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. "Brain-computer interfaces for communication and control". In: *Clinical Neurophysiology* 113.6 (June 2002), pp. 767–791.

[8] J. d. R. Millán et al. "Combining Brain–Computer Interfaces and Assistive Technologies: State-of-the-Art and Challenges". In: *Frontiers in Neuroscience* 4 (2010), p. 161.

[9] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger. "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update". en. In: *Journal of Neural Engineering* 15.3 (2018), p. 031005.

[10] K. K. Ang and C. Guan. "Brain-computer interface in stroke rehabilitation". In: *Journal of Computing Science and Engineering* 7.2 (2013), pp. 139–146.

[11] S. R. Soekadar, S. Silvoni, L. G. Cohen, and N. Birbaumer. "Brain–Machine Interfaces in Stroke Neurorehabilitation". In: *Clinical Systems Neuroscience*. Springer, 2015, pp. 3–14.

[12] A. Remsik et al. "A review of the progression and future implications of brain-computer interface therapies for restoration of distal upper extremity motor function after stroke". In: *Expert Review of Medical Devices* 13.5 (May 2016), pp. 445–454.

[13] R. Carvalho, N. Dias, and J. J. Cerqueira. "Brain-machine interface of upper limb recovery in stroke patients rehabilitation: A systematic review". en. In: *Physiotherapy Research International* (Jan. 2019), e1764.

[14] R. Sitaram et al. "Closed-loop brain training: the science of neurofeedback". en. In: *Nature Reviews Neuroscience* 18.2 (Feb. 2017), pp. 86–100.

[15] A. Ramos-Murguialday et al. "Brain–machine interface in chronic stroke rehabilitation: a controlled study". In: *Annals of Neurology* 74.1 (2013), pp. 100–108.

[16] A. Biasiucci et al. "Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke". en. In: *Nature Communications* 9.1 (June 2018), p. 2421.

[17] M. A. Cervera, S. R. Soekadar, J. Ushiba, J. d. R. Millán, M. Liu, N. Birbaumer, and G. Garipelli. "Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis". en. In: *Annals of Clinical and Translational Neurology* 5.5 (May 2018), pp. 651–663.

[18] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Z. Zhang, and K. Kreutz-Delgado. "Evolving signal processing for brain–computer interfaces". In: *Proceedings of the IEEE* 100.Special Centennial Issue (2012), pp. 1567–1584.

[19] K.-R. Müller, M. Tangermann, G. Dornhege, M. Krauledat, G. Curio, and B. Blankertz. "Machine learning for real-time single-trial EEG-analysis: from brain–computer interfacing to mental state monitoring". In: *Journal of Neuroscience Methods* 167.1 (2008), pp. 82–90.

[20] O. Jensen, A. Bahramisharif, R. Oostenveld, S. Klanke, A. Hadjipapas, Y. O. Okazaki, and M. A. van Gerven. "Using brain–computer interfaces and brain-state dependent stimulation as tools in cognitive neuroscience". In: *Frontiers in Psychology* 2 (2011), p. 100.

[21] J. M. Horschig, J. M. Zumer, and A. Bahramisharif. "Hypothesis-driven methods to augment human cognition by optimizing cortical oscillations". In: *Frontiers in Systems Neuroscience* 8 (2014).

[22] N. Lodha, S. K. Naik, S. A. Coombes, and J. H. Cauraugh. "Force control and degree of motor impairments in chronic stroke". In: *Clinical Neurophysiology* 121.11 (2010), pp. 1952 –1961.

[23] R. Ranganathan, R. Gebara, M. Andary, and J. Sylvain. "Chronic stroke survivors show task-dependent modulation of motor variability during bimanual coordination". In: *Journal of Neurophysiology* (Jan. 2019).

[24] R. Osu, K.-i. Morishige, J. Nakanishi, H. Miyamoto, and M. Kawato. "Practice reduces task relevant variance modulation and forms nominal trajectory". In: *Scientific Reports* 5 (2015).

[25] A. A. Faisal, L. P. Selen, and D. M. Wolpert. "Noise in the nervous system". In: *Nature Reviews Neuroscience* 9.4 (2008), pp. 292–303.

[26] H. G. Wu, Y. R. Miyamoto, L. N. G. Castro, B. P. Ölveczky, and M. A. Smith. "Temporal structure of motor variability is dynamically regulated and predicts motor learning ability". In: *Nature Neuroscience* 17.2 (2014), pp. 312–321.

[27] W. Singer. "Neuronal oscillations: unavoidable and useful?" en. In: *European Journal of Neuroscience* 48.7 (Oct. 2018), pp. 2389–2398.

[28] A. Meinel, S. Castaño-Candamil, J. Reis, and M. Tangermann. "Pre-Trial EEG-Based Single-Trial Motor Performance Prediction to Enhance Neuroergonomics for a Hand Force Task". In: *Frontiers in Human Neuroscience* 10 (2016), p. 170.

[29] A. Meinel, J. S. Castaño-Candamil, S. Dähne, J. Reis, and M. Tangermann. "EEG Band Power Predicts Single-Trial Reaction Time in a Hand Motor Task". In: *Proc. Int. IEEE Conf. on Neural Eng. (NER)*. Montpellier, France: IEEE, 2015, pp. 182–185.

[30] M. Tangermann, J. Reis, and A. Meinel. "Commonalities of Motor Performance Metrics are Revealed by Predictive Oscillatory EEG Components". In: *Proc. of the 3rd Int. Congr. on Neurotechnology, Electronics and Informatics (NEUROTECHNIX 2015)*. 2015, pp. 32–38.

[31] J. S. Castaño-Candamil, A. Meinel, S. Dähne, and M. Tangermann. "Probing Meaningfulness of Oscillatory EEG Components with Bootstrapping, Label Noise and Reduced Training Sets". In: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE. 2015, pp. 5159–5162.

[32] A. Meinel, J. S. Castaño-Candamil, B. Blankertz, F. Lotte, and M. Tangermann. "Characterizing Regularization Techniques for Spatial Filter Optimization in Oscillatory EEG Regression Problems". In: *Neuroinformatics* 17.2 (2019), pp. 235–251.

[33] A. Meinel, F. Lotte, and M. Tangermann. "Tikhonov Regularization Enhances EEG-Based Spatial Filtering For Single-Trial Regression". In: *Proceedings of the 7th Graz Brain-Computer interface Conference 2017*. 2017, pp. 308–313.

[34] A. Meinel, H. Kolkhorst, and M. Tangermann. "Mining within-trial oscillatory brain dynamics to address the variability of optimized spatial filters". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.3 (2019), pp. 378–388.

[35] T. J. Sejnowski, P. S. Churchland, and J. A. Movshon. *Putting big data to good use in neuroscience*. en. Comments and Opinion. Oct. 2014.

[36] H. Berger. "Uber das Elektrenkephalogramm des Menschen". de. In: *Archiv für Psychiatrie und Nervenkrankheiten* 87.1 (Dec. 1929), pp. 527–570.

[37] S. Baillet, J. C. Mosher, and R. M. Leahy. "Electromagnetic Brain Mapping". In: *Signal Processing Magazine, IEEE* 18.November (2001), pp. 14 –30.

[38] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, and A. J. Hudspeth. *Principles of neural science*. Vol. 4. McGraw-hill New York, 2000.

[39] P. L. Nunez and R. Srinivasan. *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, 2006.

[40] H. Hallez et al. "Review on solving the forward problem in EEG source analysis". In: *Journal of NeuroEngineering and Rehabilitation* 4 (Nov. 2007), p. 46.

[41] G. Buzsáki. *Rhythms of the Brain*. Oxford University Press, 2006.

[42] M. M. N. Mannan, M. A. Kamran, and M. Y. Jeong. "Identification and Removal of Physiological Artifacts From Electroencephalogram Signals: A Review". In: *IEEE Access* 6 (2018), pp. 30630–30652.

[43] F. van Ede, A. J. Quinn, M. W. Woolrich, and A. C. Nobre. "Neural Oscillations: Sustained Rhythms or Transient Burst-Events?" In: *Trends in Neurosciences* 41.7 (July 2018), pp. 415–417.

[44] W. Klimesch. "The frequency architecture of brain and brain body oscillations: an analysis". In: *European Journal of Neuroscience* 48.7 (Oct. 2018), pp. 2431–2453.

[45] W. Klimesch. "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis". In: *Brain Research Reviews* 29.2 (1999), pp. 169–195.

[46] H. van Dijk, J.-M. Schoffelen, R. Oostenveld, and O. Jensen. "Prestimulus Oscillatory Activity in the Alpha Band Predicts Visual Discrimination Ability". In: *The Journal of Neuroscience* 28.8 (2008), pp. 1816–1823.

[47] N. A. Busch, J. Dubois, and R. VanRullen. "The phase of ongoing EEG oscillations predicts visual perception". In: *The Journal of Neuroscience* 29.24 (2009), pp. 7869–7876.

[48] O. Jensen and A. Mazaheri. "Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition". English. In: *Frontiers in Human Neuroscience* 4 (2010).

[49] G. Pfurtscheller. "Graphical display and statistical evaluation of event-related desynchronization (ERD)". English. In: *Electroencephalography and Clinical Neurophysiology* 43.5 (Nov. 1977), pp. 757–760.

[50] G. Pfurtscheller and F. L. Da Silva. "Event-related EEG/MEG synchronization and desynchronization: basic principles". In: *Clinical Neurophysiology* 110.11 (1999), pp. 1842–1857.

[51] C. Neuper and G. Pfurtscheller. "Event-related dynamics of cortical rhythms: frequency-specific features and functional correlates". In: *International Journal of Psychophysiology*. Thalamo-Cortical Relationships 43.1 (Dec. 2001), pp. 41–58.

[52] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw. "Spatial filter selection for EEG-based communication". In: *Electroencephalography and Clinical Neurophysiology* 103.3 (Sept. 1997), pp. 386–394.

[53] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller. "Optimizing spatial filters for robust EEG single-trial analysis". In: *Signal Processing Magazine, IEEE* 25.1 (2008), pp. 41–56.

[54] B. Graimann and G. Pfurtscheller. "Quantification and visualization of event-related changes in oscillatory brain activity in the time-frequency domain." In: *Progress in brain research* 159 (2006), p. 79.

[55] S. Lemm, K.-R. Müller, and G. Curio. "A Generalized Framework for Quantifying the Dynamics of EEG Event-Related Desynchronization". In: *PLoS Computational Biology* 5.8 (Aug. 2009).

[56] S. Makeig. "Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones". In: *Electroencephalography and Clinical Neurophysiology* 86.4 (Apr. 1993), pp. 283–293.

[57] S. Makeig, S. Debener, J. Onton, and A. Delorme. "Mining event-related brain dynamics". In: *Trends in Cognitive Sciences* 8.5 (May 2004), pp. 204–210.

[58] M. Seeber, R. Scherer, and G. R. Müller-Putz. "EEG Oscillations Are Modulated in Different Behavior-Related Networks during Rhythmic Finger Movements". en. In: *Journal of Neuroscience* 36.46 (Nov. 2016), pp. 11671–11681.

[59] J. Onton and S. Makeig. "Information-based modeling of event-related brain dynamics". In: *Progress in Brain Research*. Ed. by C. Neuper and W. Klimesch. Vol. 159. Event-Related Dynamics of Brain Oscillations. Elsevier, Jan. 2006, pp. 99–120.

[60] A. d. Cheveigné and D. Arzounian. "Scanning for oscillations". en. In: *Journal of Neural Engineering* 12.6 (2015), p. 066020.

[61] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller. "Introduction to machine learning for brain imaging". In: *NeuroImage* 56.2 (2011), pp. 387–399.

[62] F. Lotte. "Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-Based Brain-Computer Interfaces". In: *Proceedings of the IEEE* 103.6 (June 2015), pp. 871–890.

[63] N. Birbaumer, A. R. Murguialday, and L. Cohen. "Brain-computer interface in paralysis". en-US. In: *Current Opinion in Neurology* 21.6 (Dec. 2008), p. 634.

[64] D. J. McFarland and J. R. Wolpaw. "Brain-computer interface use is a skill that user and system acquire together". en. In: *PLOS Biology* 16.7 (July 2018), e2006719.

[65] F. Nijboer et al. "A P300-based brain–computer interface for people with amyotrophic lateral sclerosis". In: *Clinical Neurophysiology* 119.8 (Aug. 2008), pp. 1909–1916.

[66] J. Höhne, E. Holz, P. Staiger-Sälzer, K.-R. Müller, A. Kübler, and M. Tangermann. "Motor Imagery for Severely Motor-Impaired Patients: Evidence for Brain-Computer Interfacing as Superior Control Solution". In: *PLoS ONE* 9.8 (Aug. 2014), e104854.

[67] S. R. Soekadar, N. Birbaumer, M. W. Slutzky, and L. G. Cohen. "Brain-machine interfaces in neurorehabilitation of stroke". In: *Neurobiology of Disease* 83 (Nov. 2015), pp. 172–179.

[68] L. Tonin, M. Pitteri, R. Leeb, H. Zhang, E. Menegatti, F. Piccione, and J. d. R. Millán. "Behavioral and Cortical Effects during Attention Driven Brain-Computer Interface Operations in Spatial Neglect: A Feasibility Case Study". English. In: *Frontiers in Human Neuroscience* 11 (2017).

[69] L. Carelli et al. *Brain-Computer Interface for Clinical Purposes: Cognitive Assessment and Rehabilitation*. en. Research article. 2017.

[70] M. Musso, D. Hübner, S. Schwarzkopf, C. Weiller, and M. Tangermann. "Evidence for language-specificity of a BCI-based language training". In: 69. 2018.

[71] B. M. Young et al. "Changes in functional connectivity correlate with behavioral gains in stroke patients after therapy using a brain-computer interface device". English. In: *Frontiers in Neuroengineering* 7 (2014).

[72] K. K. Ang et al. "Brain-computer interface-based robotic end effector system for wrist and hand rehabilitation: results of a three-armed randomized controlled trial for chronic stroke". English. In: *Frontiers in Neuroengineering* 7 (2014).

[73] A. A. Frolov et al. "Post-stroke Rehabilitation Training with a Motor-Imagery-Based Brain-Computer Interface (BCI)-Controlled Hand Exoskeleton: A Randomized Controlled Multicenter Trial". English. In: *Frontiers in Neuroscience* 11 (2017).

[74] F. Pichiorri et al. "Brain–computer interface boosts motor imagery practice during stroke recovery". In: *Annals of Neurology* 77.5 (2015), pp. 851–865.

[75] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda. "Recipes for the linear analysis of EEG". In: *NeuroImage* 28.2 (2005), pp. 326–341.

[76] S. Haufe, F. Meinecke, K. Görgen, S. Dähne, J.-D. Haynes, B. Blankertz, and F. Bießmann. "On the interpretation of weight vectors of linear models in multivariate neuroimaging". In: *NeuroImage* 87.0 (2014), pp. 96 –110.

[77] D. A. Bridwell, J. F. Cavanagh, A. G. E. Collins, M. D. Nunez, R. Srinivasan, S. Stober, and V. D. Calhoun. "Moving Beyond ERP Components: A Selective Review of Approaches to Integrate EEG and Behavior". English. In: *Frontiers in Human Neuroscience* 12 (2018).

[78] G. H. Golub and C. F. V. Loan. *Matrix Computations*. en. JHU Press, Dec. 2012.

[79] S. Dähne et al. "Multivariate Machine Learning Methods for Fusing Multimodal Functional Neuroimaging Data". In: *Proceedings of the IEEE* 103.9 (Sept. 2015), pp. 1507–1530.

[80] A. de Cheveigné and L. C. Parra. "Joint decorrelation, a versatile tool for multichannel data analysis". In: *NeuroImage* 98 (Sept. 2014), pp. 487–505.

[81] H. Hotelling. "Analysis of a complex of statistical variables into principal components". In: *Journal of Educational Psychology* 24.7 (1933), pp. 498–520.

[82] S. Haufe, S. Dähne, and V. V. Nikulin. "Dimensionality reduction for the analysis of brain oscillations". In: *NeuroImage* 101 (Nov. 2014), pp. 583–597.

[83] R. K. Niazy, C. F. Beckmann, G. D. Iannetti, J. M. Brady, and S. M. Smith. "Removal of FMRI environment artifacts from EEG data using optimal basis sets". eng. In: *NeuroImage* 28.3 (Nov. 2005), pp. 720–737.

[84] A. J. Bell and T. J. Sejnowski. "An Information-Maximization Approach to Blind Separation and Blind Deconvolution". In: *Neural Computation* 7.6 (Nov. 1995), pp. 1129–1159.

[85] A. Hyvarinen. "Fast and robust fixed-point algorithms for independent component analysis". In: *IEEE Transactions on Neural Networks* 10.3 (May 1999), pp. 626–634.

[86] A. Belouchrani, K. Abed-Meraim, J. Cardoso, and E. Moulines. "A blind source separation technique using second-order statistics". In: *IEEE Transactions on Signal Processing* 45.2 (Feb. 1997), pp. 434–444.

[87] V. D. Calhoun, J. Liu, and T. Adali. "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data". In: *NeuroImage*. Mathematics in Brain Imaging 45.1, Supplement 1 (Mar. 2009), S163–S172.

[88] C. J. James and C. W. Hesse. "Independent component analysis for biomedical signals". en. In: *Physiological Measurement* 26.1 (2005), R15.

[89] I. Winkler, S. Brandl, F. Horn, E. Waldburger, C. Allefeld, and M. Tangermann. "Robust artifactual independent component classification for BCI practitioners". In: *Journal of Neural Engineering* 11.3 (2014), p. 035013.

[90] F. Artoni, D. Menicucci, A. Delorme, S. Makeig, and S. Micera. "RELICA: A method for estimating the reliability of independent components". In: *NeuroImage* 103 (Dec. 2014), pp. 391–400.

[91] Z. J. Koles. "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG". In: *Electroencephalography and clinical Neurophysiology* 79.6 (1991), pp. 440–447.

[92] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller. "Optimal spatial filtering of single trial EEG during imagined hand movement". In: *Rehabilitation Engineering, IEEE Transactions on* 8.4 (2000), pp. 441–446.

[93] F. Lotte and C. Guan. "Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms". In: *IEEE Transactions on Biomedical Engineering* 58.2 (2011), pp. 355–362.

[94] S. Dähne, F. C. Meinecke, S. Haufe, J. Höhne, M. Tangermann, K.-R. Müller, and V. V. Nikulin. "SPoC: a novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters". In: *NeuroImage* 86.0 (2014), pp. 111–122.

[95] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller. "Spatio-spectral filters for improving the classification of single trial EEG". In: *IEEE Transactions on Biomedical Engineering* 52.9 (Sept. 2005), pp. 1541–1548.

[96] A. Gramfort, D. Strohmeier, J. Haueisen, M. Hämäläinen, and M. Kowalski. "Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations". In: *NeuroImage* 70 (2013), pp. 410 –422.

[97] R. Grech et al. "Review on solving the inverse problem in EEG source analysis". In: 5.25 (2008), pp. 792 –800.

[98] S. Castaño-Candamil, J. Höhne, J.-D. Martínez-Vargas, X.-W. An, G. Castellanos-Domínguez, and S. Haufe. "Solving the EEG inverse problem based on space-time-frequency structured sparsity constraints". In: *NeuroImage* 118 (2015), pp. 598 –612.

[99] Y. Huang, L. C. Parra, and S. Haufe. "The New York Head-A precise standardized volume conductor model for EEG source localization and tES targeting". In: *NeuroImage* 140 (2016). Transcranial electric stimulation (tES) and Neuroimaging, pp. 150 –162.

[100] R. T. Schirrmeister et al. "Deep learning with convolutional neural networks for EEG decoding and visualization". en. In: *Human Brain Mapping* 38.11 (2017), pp. 5391–5420.

[101] G. Montavon, W. Samek, and K.-R. Müller. "Methods for interpreting and understanding deep neural networks". In: *Digital Signal Processing* 73 (Feb. 2018), pp. 1–15.

[102] J. W. Krakauer. "Motor learning: its relevance to stroke recovery and neurorehabilitation". In: *Current opinion in Neurology* 19.1 (2006), pp. 84–90.

[103] P. Langhorne, J. Bernhardt, and G. Kwakkel. "Stroke rehabilitation". In: *The Lancet* 377.9778 (2011), pp. 1693–1702.

[104] B. H. Dobkin. "Strategies for stroke rehabilitation". In: *The Lancet Neurology* 3.9 (2004), pp. 528–536.

[105] A. A. Timmermans, H. A. Seelen, R. D. Willmann, and H. Kingma. "Technology-assisted training of arm-hand skills in stroke: concepts on reacquisition of motor control and therapist guidelines for rehabilitation technology design". In: *Journal of Neuroengineering and Rehabilitation* 6.1 (2009), p. 1.

[106] B. French et al. "Repetitive task training for improving functional ability after stroke". In: *Cochrane Database Syst Rev* 4.4 (2007).

[107] S. L. Wolf, S. Blanton, H. Baer, J. Breshears, and A. J. Butler. "Repetitive task practice: a critical review of constraint-induced movement therapy in stroke". In: *The Neurologist* 8.6 (2002), p. 325.

[108] F. Coupar, A. Pollock, F. Van Wijck, J. Morris, and P. Langhorne. "Simultaneous bilateral training for improving arm function after stroke". In: *Cochrane Database Syst Rev* 4.4 (2010).

[109] G. Kwakkel, B. J. Kollen, and H. I. Krebs. "Effects of robot-assisted therapy on upper limb recovery after stroke: a systematic review". In: *Neurorehabilitation and Neural Repair* (2007).

[110] M. O. Abe and D. Sternad. "Directionality in Distribution and Temporal Structure of Variability in Skill Acquisition". In: *Frontiers in Human Neuroscience* 7.225 (2013).

[111] A. M. Hadjiosif and M. A. Smith. "Flexible control of safety margins for action based on environmental variability". In: *The Journal of Neuroscience* 35.24 (2015), pp. 9106–9121.

[112] R. J. van Beers, Y. van der Meer, and R. M. Veerman. "What autocorrelation tells us about motor variability: Insights from dart throwing". In: *PloS one* 8.5 (2013), e64332.

[113] G. M. Lage, H. Ugrinowitsch, T. Apolinário-Souza, M. M. Vieira, M. R. Albuquerque, and R. N. Benda. "Repetition and variation in motor practice: A review of neural correlates". In: *Neuroscience & Biobehavioral Reviews* 57 (2015), pp. 132–141.

[114] R. Cohen and D. Sternad. "Variability in motor learning: relocating, channeling and reducing noise". In: *Experimental Brain Research* 193.1 (2009), pp. 69–83.

[115] J. Reis et al. "Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation". In: *Proceedings of the National Academy of Sciences* (2009).

[116] B. Fritsch, J. Reis, K. Martinowich, H. M. Schambra, Y. Ji, L. G. Cohen, and B. Lu. "Direct current stimulation promotes BDNF-dependent synaptic plasticity: potential implications for motor learning". In: *Neuron* 66.2 (2010), pp. 198–204.

[117] M. Camus, P. Ragert, Y. Vandermeeren, and L. G. Cohen. "Mechanisms controlling motor output to a transfer hand after learning a sequential pinch force skill with the opposite hand". In: *Clinical Neurophysiology* 120.10 (2009), pp. 1859–1865.

[118] V. S. Handiru, A. P. Vinod, and C. Guan. "EEG Source Imaging of Movement Decoding: The State of the Art and Future Directions". In: *IEEE Systems, Man, and Cybernetics Magazine* 4.2 (Apr. 2018), pp. 14–23.

[119] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio. "The non-invasive Berlin brain–computer interface: fast acquisition of effective performance in untrained subjects". In: *NeuroImage* 37.2 (2007), pp. 539–550.

[120] P.-J. Kindermans, M. Tangermann, K.-R. Müller, and B. Schrauwen. "Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP speller". In: *Journal of Neural Engineering* 11.3 (2014), p. 035005.

[121] S. Hanslmayr, A. Aslan, T. Staudigl, W. Klimesch, C. S. Herrmann, and K.-H. Bäuml. "Prestimulus oscillations predict visual perception performance between and within subjects". In: *NeuroImage* 37.4 (2007), pp. 1465–1473.

[122] L. Tonin, R. Leeb, A. Sobolewski, and J del R Millán. "An online EEG BCI based on covert visuospatial attention in absence of exogenous stimulation". In: *Journal of Neural Engineering* 10.5 (2013), p. 056007.

[123] J. M. Horschig, W. Oosterheert, R. Oostenveld, and O. Jensen. "Modulation of Posterior Alpha Activity by Spatial Attention Allows for Controlling A Continuous Brain–Computer Interface". In: *Brain Topography* (2014), pp. 1–13.

[124] S. L. Gonzalez Andino, C. M. Michel, G. Thut, T. Landis, and R. Grave de Peralta. "Prediction of response speed by anticipatory high-frequency (gamma band) oscillations in the human brain". In: *Human Brain Mapping* 24.1 (2005), pp. 50–58.

[125] N. Hoogenboom, J.-M. Schoffelen, R. Oostenveld, and P. Fries. "Visually induced gamma-band activity predicts speed of change detection in humans". In: *NeuroImage* 51.3 (2010), pp. 1162–1167.

[126] R. Schubert, S. Haufe, F. Blankenburg, A. Villringer, and G. Curio. "Now you'll feel it, now you won't: EEG rhythms predict the effectiveness of perceptual masking". In: *Journal of Cognitive Neuroscience* 21.12 (2009), pp. 2407–2419.

[127] A. P. Georgopoulos, R. E. Kettner, and A. B. Schwartz. "Primate motor cortex and free arm movements to visual targets in three- dimensional space. II. Coding of the direction of movement by a neuronal population". In: *Journal of Neuroscience* 8.8 (Aug. 1988), pp. 2928–2937.

[128] K. Jerbi et al. "Inferring hand movement kinematics from MEG, EEG and intracranial EEG: From brain-machine interfaces to motor rehabilitation". In: *IRBM* 32.1 (2011), pp. 8 –18.

[129] T. Pistohl, T. Ball, A. Schulze-Bonhage, A. Aertsen, and C. Mehring. "Prediction of arm movement trajectories from ECoG-recordings in humans". In: *Journal of Neuroscience Methods* 167.1 (2008), pp. 105–114.

[130] R. D. Flint, J. M. Rosenow, M. C. Tate, and M. W. Slutzky. "Continuous decoding of human grasp kinematics using epidural and subdural signals". en. In: *Journal of Neural Engineering* 14.1 (2017), p. 016005.

[131] S. Waldert, H. Preissl, E. Demandt, C. Braun, N. Birbaumer, A. Aertsen, and C. Mehring. "Hand movement direction decoded from MEG and EEG". In: *The Journal of Neuroscience* 28.4 (2008), pp. 1000–1008.

[132] T. J. Bradberry, F. Rong, and J. L. Contreras-Vidal. "Decoding center-out hand velocity from MEG signals during visuomotor adaptation". In: *NeuroImage* 47.4 (2009), pp. 1691–1700.

[133] V. Logar, I. Škrjanc, A. Belič, R. Karba, S. Brežan, B. Koritnik, and J. Zidar. "Gripping-force identification using EEG and phase-demodulation approach". In: *Neuroscience Research* 60.4 (2008), pp. 389–396.

[134] M. Zaepffel, R. Trachel, B. E. Kilavik, and T. Brochier. "Modulations of EEG beta power during planning and execution of grasping movements". In: *PloS One* 8.3 (2013), e60060.

[135] E. Y. Lew, R. Chavarriaga, S. Silvoni, and J. d. R. Millán. "Single trial prediction of self-paced reaching directions from EEG signals". In: *Frontiers in Neuroscience* 8 (2014).

[136] P. S. Hammon, S. Makeig, H. Poizner, E. Todorov, and V. D. Sa. "Predicting reaching targets from human EEG". In: *Signal Processing Magazine, IEEE* 25.1 (2008), pp. 69–77.

[137] C. L. Maeder, C. Sannelli, S. Haufe, and B. Blankertz. "Pre-stimulus sensorimotor rhythms influence brain–computer interface classification performance". In: *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 20.5 (2012), pp. 653–662.

[138] L. Yang, H. Leung, M. Plank, J. Snider, and H. Poizner. "Alpha and beta band power changes predict reaction time and endpoint error during planning reaching movements". In: *Biomedical Engineering and Informatics (BMEI), 2014 7th International Conference on*. IEEE. 2014, pp. 264–268.

[139] T. Meyer, J. Peters, T. O. Zander, B. Schölkopf, and M. Grosse-Wentrup. "Predicting motor learning performance from Electroencephalographic data". In: *Journal of NeuroEngineering and Rehabilitation* 11.1 (2014), p. 24.

[140] B. Ovbiagele and M. N. Nguyen-Huynh. "Stroke epidemiology: advancing our understanding of disease mechanism and therapy". In: *Neurotherapeutics* 8.3 (2011), pp. 319–329.

[141] T. Fawcett. "An introduction to ROC analysis". In: *Pattern Recognition Letters* 27.8 (2006), pp. 861–874.

[142] A. Meinel, E.-M. Schlichtmann, T. Koller, J. Reis, and M. Tangermann. "Predicting Single-Trial Motor Performance from Oscillatory EEG in Chronic Stroke Patients." In: *Proc. of the Int. Brain-Computer Interface Meeting: Past, Present, and Future*. 2016, p. 140.

[143] F. Artoni, A. Delorme, and S. Makeig. "Applying dimension reduction to EEG data by Principal Component Analysis reduces the quality of its subsequent Independent Component decomposition". In: *NeuroImage* 175 (July 2018), pp. 176–187.

[144] I. Daly, R. Scherer, M. Billinger, and G. Muller-Putz. "FORCe: Fully Online and automated artifact Removal for brain-Computer interfacing". In: *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 23.5 (2015), pp. 725–736.

[145] M. Grosse-Wentrup, S. Harmeling, T. Zander, J. Hill, and B. Scholkopf. "How to Test the Quality of Reconstructed Sources in Independent Component Analysis (ICA) of EEG/MEG Data". In: *Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on*. IEEE. 2013, pp. 102–105.

[146] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup. "Transfer Learning in Brain-Computer Interfaces". In: *IEEE Computational Intelligence Magazine* 11.1 (Feb. 2016), pp. 20–31.

[147] E. M. Hammer et al. "Psychological predictors of SMR-BCI performance". In: *Biological Psychology* 89.1 (2012), pp. 80–86.

[148] B. Z. Allison, C. Brunner, C. Altstätter, I. C. Wagner, S. Grissmann, and C. Neuper. "A hybrid ERD/SSVEP BCI for continuous simultaneous two dimensional cursor control". In: *Journal of Neuroscience Methods* 209.2 (2012), pp. 299 –307.

[149] G. R. Müller-Putz et al. "Towards Non-Invasive Hybrid Brain-Computer Interfaces: Framework, Practice, Clinical Application and Beyond". In: *Proceedings of the IEEE* 103.6 (2015), pp. 926–943.

[150] C. Sannelli, T. Dickhaus, S. Halder, E. M. Hammer, K.-R. Müller, and B. Blankertz. "On optimal channel configurations for SMR-based brain-computer interfaces". In: *Brain Topography* 23.2 (2010), pp. 186–193.

[151] S.-H. Hsu, T. Mullen, T.-P. Jung, and G. Cauwenberghs. "Validating online recursive independent component analysis on EEG data". In: *Neural Engineering (NER), 2015 7th International IEEE/EMBS Conference on*. IEEE. 2015, pp. 918–921.

[152] M. Tangermann and A. Meinel. "Informative Oscillatory EEG Components and their Persistence in Time and Frequency". In: *NEUROTECHNIX 2017 - Extended Abstracts - CogNeuroEng*, INSTICC. SciTePress, 2017, pp. 17–21.

[153] A. Meinel, T. Koller, and M. Tangermann. "Time-frequency sensitivity characterization of single-trial oscillatory EEG components". In: *The First Biannual Neuroadaptive Technology Conference*. 2017, pp. 36–37.

[154] F. van Ede, M. Köster, and E. Maris. "Beyond establishing involvement: Quantifying the contribution of anticipatory α-and β-band suppression to perceptual improvement with attention". In: *Journal of Neurophysiology* 108.9 (2012), pp. 2352–2362.

[155] S. Castaño-Candamil, A. Meinel, J. Reis, and M. Tangermann. "P186. Correlates to influence user performance in a hand motor rehabilitation task". In: *Clinical Neurophysiology* 126.8 (2015), e166 –e167.

[156] M. Krauledat, M. Tangermann, B. Blankertz, and K.-R. Müller. "Towards Zero Training for Brain-Computer Interfacing". In: *PLoS ONE* 3.8 (Aug. 2008), e2967.

[157] J. M. Horschig, O. Jensen, M. R. van Schouwenburg, R. Cools, and M. Bonnefond. "Alpha activity reflects individual abilities to adapt to the environment". In: *NeuroImage* 89 (2014), pp. 235–243.

[158] K. S. Chaisanguanthum, H. H. Shen, and P. N. Sabes. "Motor variability arises from a slow random walk in neural state". In: *The Journal of Neuroscience* 34.36 (2014), pp. 12071–12080.

[159] M. M. Churchland, M. Y. Byron, S. I. Ryu, G. Santhanam, and K. V. Shenoy. "Neural variability in premotor cortex provides a signature of motor preparation". In: *The Journal of neuroscience* 26.14 (2006), pp. 3697–3712.

[160] J. Farquhar and N. J. Hill. "Interactions Between Pre-Processing and Classification Methods for Event-Related-Potential Classification". In: *Neuroinformatics* 11.2 (2013), pp. 175–192.

[161] D. Krusienski, M Grosse-Wentrup, F Galán, D Coyle, K. Miller, E Forney, and C. Anderson. "Critical issues in state-of-the-art brain-computer interface signal processing". In: *Journal of Neural Engineering* 8.2 (2011), p. 025002.

[162] B. Reuderink and M. Poel. *Robustness of the Common Spatial Patterns algorithm in the BCI-pipeline*. Tech. rep. HMI, University of Twente, 2008.

[163] W. Samek, M. Kawanabe, and K.-R. Müller. "Divergence-Based Framework for Common Spatial Patterns Algorithms". In: *IEEE Reviews in Biomedical Engineering* 7 (2014), pp. 50–72.

[164] M. Grosse-Wentrup, C. Liefhold, K. Gramann, and M. Buss. "Beamforming in non invasive Brain-Computer Interfaces". In: *IEEE Transactions on Biomedical Engineering* 56.4 (2009), pp. 1209 –1219.

[165] S. H. Park, D. Lee, and S. G. Lee. "Filter Bank Regularized Common Spatial Pattern Ensemble for Small Sample Motor Imagery Classification". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* PP.99 (2017), pp. 1–1.

[166] T. S. Tian, J. Z. Huang, H. Shen, and Z. Li. "EEG/MEG Source Reconstruction with Spatial-Temporal Two-Way Regularized Regression". In: *Neuroinformatics* 11.4 (2013), pp. 477–493.

[167] H. Wang and X. Li. "Regularized Filters for L1-Norm-Based Common Spatial Patterns". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 24.2 (Feb. 2016), pp. 201–211.

[168] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek. "Optimizing the Channel Selection and Classification Accuracy in EEG-Based BCI". In: *IEEE Transactions on Biomedical Engineering* 58.6 (June 2011), pp. 1865–1873.

[169] J. Farquhar, N. Hill, Lal, T.N., and Schölkopf, B. "Regularised CSP for sensor selection in BCI". In: *Proceedings of the 3rd international BCI workshop*. 2006.

[170] M. Cheng, Z. Lu, and H. Wang. "Regularized common spatial patterns with subject-to-subject transfer of EEG signals". en. In: *Cognitive Neurodynamics* 11.2 (Apr. 2017), pp. 173–181.

[171] D. Devlaminck, B. Wyns, M. Grosse-Wentrup, G. Otte, and P. Santens. "Multisubject Learning for Common Spatial Patterns in Motor-imagery BCI". In: *Intell. Neuroscience* 2011 (Jan. 2011), 8:8–8:8.

[172] W. Samek, F. C. Meinecke, and K.-R. Müller. "Transferring Subspaces Between Subjects in Brain-Computer Interfacing". In: *IEEE Transactions on Biomedical Engineering* 60.8 (Aug. 2013), pp. 2289–2298.

[173] H. Kang, Y. Nam, and S. Choi. "Composite Common Spatial Pattern for Subject-to-Subject Transfer". In: *IEEE Signal Processing Letters* 16.8 (Aug. 2009), pp. 683–686.

[174] F. Lotte and C. Guan. "Learning from other subjects helps reducing Brain-Computer Interface calibration time". In: *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. Mar. 2010, pp. 614–617.

[175] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek. "Optimizing Spatial Filters by Minimizing Within-Class Dissimilarities in Electroencephalogram-Based Brain-Computer interface". In: *IEEE Transactions on Neural Networks and Learning Systems* 24.4 (Apr. 2013), pp. 610–619.

[176] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe. "Stationary common spatial patterns for brain-computer interfacing". en. In: *Journal of Neural Engineering* 9.2 (2012), p. 026013.

[177] H. Cho, M. Ahn, K. Kim, and S. C. Jun. "Increasing session-to-session transfer in a brain–computer interface with on-site background noise acquisition". en. In: *Journal of Neural Engineering* 12.6 (2015), p. 066009.

[178] D. Wu, J. T. King, C. H. Chuang, C. T. Lin, and T. P. Jung. "Spatial Filtering for EEG-Based Regression Problems in brain-computer interface (BCI)". In: *IEEE Transactions on Fuzzy Systems* PP.99 (2017), pp. 1–1.

[179] J. Frey, M. Daniel, M. Hachet, J. Castet, and F. Lotte. "Framework for electroence-phalography-based evaluation of user experience". In: *Procedings of CHI*. 2016, pp. 2283–2294.

[180] M. Schultze-Kraft, S. Dähne, M. Gugler, G. Curio, and B. Blankertz. "Unsupervised classification of operator workload from brain signals". en. In: *Journal of Neural Engineering* 13.3 (2016), p. 036008.

[181] I.-E. Nicolae, L. Acqualagna, and B. Blankertz. "Assessing the Depth of Cognitive Processing as the Basis for Potential User-State Adaptation". English. In: *Frontiers in Neuroscience* 11 (2017).

[182] A. Úbeda, J. M. Azorín, R. Chavarriaga, and J. d. R. Millán. "Classification of upper limb center-out reaching tasks by means of EEG-based continuous decoding techniques". In: *Journal of NeuroEngineering and Rehabilitation* 14 (Feb. 2017), p. 9.

[183] A. N. Tikhonov. "Regularization of incorrectly posed problems". In: *Soviet Mathematics Doklady* 4 (1963), pp. 1624–1627.

[184] H. Lu, H. L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos. "Regularized Common Spatial Pattern With Aggregation for EEG Classification in Small-Sample Setting". In: *IEEE Transactions on Biomedical Engineering* 57.12 (2010), pp. 2936–2946.

[185] Z. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. en. Springer Science & Business Media, Dec. 2009.

[186] O. Ledoit and M. Wolf. "A well-conditioned estimator for large-dimensional covariance matrices". In: *Journal of Multivariate Analysis* 88.2 (Feb. 2004), pp. 365–411.

[187] J. Schäfer and K. Strimmer. "A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics". In: *Statistical applications in genetics and molecular biology* 4.1 (Nov. 2005).

[188] D. Bartz and K.-R. Müller. "Covariance shrinkage for autocorrelated data". In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger. Curran Associates, Inc., 2014, pp. 1592–1600.

[189] S. Castaño-Candamil, A. Meinel, and M. Tangermann. "Post-hoc labeling of arbitrary EEG recordings for data-efficient evaluation of neural decoding methods". In: *arXiv preprint arXiv:1711.08208* (2017).

[190] B. Blankertz et al. "Neurophysiological predictor of SMR-based BCI performance". In: *NeuroImage* 51.4 (July 2010), pp. 1303–1309.

[191] T. W. Picton, M. S. John, A. Dimitrijevic, and D. Purcell. "Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos". In: *International Journal of Audiology* 42.4 (Jan. 2003), pp. 177–219.

[192] J. F. Kenney. *Mathematics of statistics*. D. Van Nostrand Company Inc; Toronto; Princeton; New Jersey; London; New York,; Affiliated East-West Press Pvt-Ltd; New Delhi, 2013.

[193] V. Jayaram and A. Barachant. "MOABB: Trustworthy algorithm benchmarking for BCIs". In: *Journal of Neural Engineering* 15.6 (Dec. 2018). arXiv: 1805.06427, p. 066011.

[194] J. Mattout, C. Phillips, W. D. Penny, M. D. Rugg, and K. J. Friston. "MEG source localization under multiple constraints: An extended Bayesian framework". In: *NeuroImage* 30.3 (Apr. 2006), pp. 753–767.

[195] Y. Chen, A. Wiesel, and A. O. Hero. "Robust Shrinkage Estimation of High-Dimensional Covariance Matrices". In: *IEEE Transactions on Signal Processing* 59.9 (Sept. 2011), pp. 4097–4107.

[196] D. A. Engemann and A. Gramfort. "Automated model selection in covariance estimation and spatial whitening of MEG and EEG signals". In: *NeuroImage* 108 (Mar. 2015), pp. 328–342.

[197] A. Meinel, K. Eggensperger, M. Tangermann, and F. Hutter. "Hyperparameter Optimization for Machine Learning Problems in BCI". In: *Proc. of the Int. Brain-Computer Interface Meeting: Past, Present, and Future*. 2016, p. 184.

[198] S. Debener, M. Ullsperger, M. Siegel, and A. K. Engel. "Single-trial EEG–fMRI reveals the dynamics of cognitive function". In: *Trends in Cognitive Sciences* 10.12 (Dec. 2006), pp. 558–563.

[199] S. Harmeling, G. Dornhege, D. Tax, F. Meinecke, and K.-R. Müller. "From outliers to prototypes: Ordering data". In: *Neurocomputing*. Blind Source Separation and Independent Component Analysis 69.13 (Aug. 2006), pp. 1608–1618.

[200] T. M. Rutkowski, D. P. Mandic, A. Cichocki, and A. W. Przybyszewski. "Emd approach to multichannel eeg data — the amplitude and phase components clustering analysis". In: *Journal of Circuits, Systems and Computers* 19.01 (Feb. 2010), pp. 215–229.

[201] S. Spadone, F. de Pasquale, D. Mantini, and S. Della Penna. "A K-means multivariate approach for clustering independent components from magnetoencephalographic data". In: *NeuroImage* 62.3 (Sept. 2012), pp. 1912–1923.

[202] N. Bigdely-Shamlo, T. Mullen, K. Kreutz-Delgado, and S. Makeig. "Measure projection analysis: A probabilistic approach to EEG source comparison and multi-subject inference". In: *NeuroImage* 72 (May 2013), pp. 287–303.

[203] J. Touryan, B. J. Lance, S. E. Kerick, A. J. Ries, and K. McDowell. "Common EEG features for behavioral estimation in disparate, real-world tasks". In: *Biological Psychology* 114 (Feb. 2016), pp. 93–107.

[204] F. Rong and J. L. Contreras-Vidal. "Magnetoencephalographic artifact identification and automatic removal based on independent component analysis and categorization approaches". In: *Journal of Neuroscience Methods* 157.2 (Oct. 2006), pp. 337–354.

[205] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[206] T. Warren Liao. "Clustering of time series data—a survey". In: *Pattern Recognition* 38.11 (Nov. 2005), pp. 1857–1874.

[207] A. K. Jain. "Data clustering: 50 years beyond K-means". In: *Pattern Recognition Letters*. Award winning papers from the 19th International Conference on Pattern Recognition (ICPR) 31.8 (June 2010), pp. 651–666.

[208] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. "A density-based algorithm for discovering clusters in large spatial databases with noise." In: *Kdd*. Vol. 96. 1996, pp. 226–231.

[209] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu. "DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN". In: *ACM Trans. Database Syst.* 42.3 (July 2017), 19:1–19:21.

[210] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu. "Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications". en. In: *Data Mining and Knowledge Discovery* 2.2 (June 1998), pp. 169–194.

[211] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona. "An extensive comparative study of cluster validity indices". In: *Pattern Recognition* 46.1 (2013), pp. 243–256.

[212] J. MacQueen. "Some methods for classification and analysis of multivariate observations". EN. In: The Regents of the University of California, 1967.

[213] D. Arthur and S. Vassilvitskii. "k-means++: The advantages of careful seeding". In: *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[214] M. Halkidi and M. Vazirgiannis. "Clustering validity assessment: finding the optimal partitioning of a data set". In: *Proceedings 2001 IEEE International Conference on Data Mining*. 2001, pp. 187–194.

[215] D. Moulavi, P. Jaskowiak, R. Campello, A. Zimek, and J. Sander. "Density-Based Clustering Validation". In: *Proceedings of the 2014 SIAM International Conference on Data Mining*. Proceedings. Society for Industrial and Applied Mathematics, Apr. 2014, pp. 839–847.

[216] P. J. Rousseeuw. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis". In: *Journal of computational and applied mathematics* 20 (1987), pp. 53–65.

[217] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu. "Understanding of Internal Clustering Validation Measures". In: *2010 IEEE International Conference on Data Mining*. Dec. 2010, pp. 911–916.

[218] S. Mehrkanoon, T. W. Boonstra, M. Breakspear, M. Hinder, and J. J. Summers. "Upregulation of cortico-cerebellar functional connectivity after motor learning". In: *NeuroImage* 128 (Mar. 2016), pp. 252–263.

[219] K. Mahjoory, V. V. Nikulin, L. Botrel, K. Linkenkaer-Hansen, M. M. Fato, and S. Haufe. "Consistency of EEG source localization and connectivity estimates". In: *NeuroImage* 152 (May 2017), pp. 590–601.

[220] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. H. Lopes da Silva. "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks". In: *NeuroImage* 31.1 (May 2006), pp. 153–159.

[221] C. C. Aggarwal, A. Hinneburg, and D. A. Keim. "On the Surprising Behavior of Distance Metrics in High Dimensional Space". en. In: *Database Theory — ICDT 2001*. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, Jan. 2001, pp. 420–434.

[222] K. C. Dodd, V. A. Nair, and V. Prabhakaran. "Role of the Contralesional vs. Ipsilesional Hemisphere in Stroke Recovery". English. In: *Frontiers in Human Neuroscience* 11 (2017).

[223] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller. "Single-trial analysis and classification of ERP components — a tutorial". In: *NeuroImage* 56.2 (2011), pp. 814–825.

[224] J.-D. Haynes and G. Rees. "Neuroimaging: Decoding mental states from brain activity in humans". en. In: *Nature Reviews Neuroscience* 7.7 (July 2006), pp. 523–534.

[225] O. Cohen, E. Sherman, N. Zinger, S. Perlmutter, and Y. Prut. "Getting ready to move: transmitted information in the corticospinal pathway during preparation for movement". In: *Current Opinion in Neurobiology*. Motor systems – Neurobiology of behaviourv 20.6 (Dec. 2010), pp. 696–703.

[226] D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw. "Effects of training pre-movement sensorimotor rhythms on behavioral performance". In: *Journal of neural engineering* 12.6 (2015), p. 066021.

[227] S. L. Norman, D. J. McFarland, A. Miner, S. C. Cramer, E. T. Wolbrecht, J. R. Wolpaw, and D. J. Reinkensmeyer. "Controlling pre-movement sensorimotor rhythm can improve finger extension after stroke". en. In: *Journal of Neural Engineering* 15.5 (2018), p. 056026.

[228] M. Hamoudi, H. M. Schambra, B. Fritsch, A. Schoechlin-Marx, C. Weiller, L. G. Cohen, and J. Reis. "Transcranial Direct Current Stimulation Enhances Motor Skill Learning but Not Generalization in Chronic Stroke". In: *Neurorehabilitation and Neural Repair* (Apr. 2018).

[229] D. J. Gladstone, C. J. Danells, and S. E. Black. "The Fugl-Meyer Assessment of Motor Recovery after Stroke: A Critical Review of Its Measurement Properties". en. In: *Neurorehabilitation and Neural Repair* 16.3 (Sept. 2002), pp. 232–240.

[230] A. R. Fugl-Meyer, L. Jääskö, I. Leyman, S. Olsson, and S. Steglind. "The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance." eng. In: *Scandinavian journal of rehabilitation medicine* 7.1 (1975), pp. 13–31.

[231] M. Jas, D. A. Engemann, Y. Bekhti, F. Raimondo, and A. Gramfort. "Autoreject: Automated artifact rejection for MEG and EEG data". In: *NeuroImage* 159 (Oct. 2017), pp. 417–429.

[232] D. J. Krusienski, M. Grosse-Wentrup, F. Galán, D. Coyle, K. J. Miller, Elliott Forney, and C. W. Anderson. "Critical issues in state-of-the-art brain–computer interface signal processing". en. In: *Journal of Neural Engineering* 8.2 (2011), p. 025002.

[233] H. Shimodaira. "Improving predictive inference under covariate shift by weighting the log-likelihood function". In: *Journal of Statistical Planning and Inference* 90.2 (Oct. 2000), pp. 227–244.

[234] H. Raza, H. Cecotti, Y. Li, and G. Prasad. "Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface". en. In: *Soft Computing* 20.8 (Aug. 2016), pp. 3085–3096.

[235] M. Spüler, W. Rosenstiel, and M. Bogdan. "Principal component based covariate shift adaption to reduce non-stationarity in a MEG-based brain-computer interface". In: *EURASIP Journal on Advances in Signal Processing* 2012.1 (July 2012), p. 129.

[236] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz. "Co-adaptive calibration to improve BCI efficiency". en. In: *Journal of Neural Engineering* 8.2 (2011), p. 025009.

[237] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller. "Towards adaptive classification for BCI". en. In: *Journal of Neural Engineering* 3.1 (2006), R13.

[238] C. Vidaurre, M. Kawanabe, P. von Bünau, B. Blankertz, and K.-R. Müller. "Toward unsupervised adaptation of LDA for brain-computer interfaces". eng. In: *IEEE transactions on bio-medical engineering* 58.3 (Mar. 2011), pp. 587–597.

[239] I. Winkler, S. Debener, K.-R. Müller, and M. Tangermann. "On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Aug. 2015, pp. 4101–4105.

[240] C. Neuper, M. Wörtz, and G. Pfurtscheller. "ERD/ERS patterns reflecting sensorimotor activation and deactivation". In: *Progress in Brain Research*. Ed. by C. Neuper and W. Klimesch. Vol. 159. Event-Related Dynamics of Brain Oscillations. Elsevier, Jan. 2006, pp. 211–222.

[241] S. J. Page, G. D. Fulk, and P. Boyne. "Clinically Important Differences for the Upper-Extremity Fugl-Meyer Scale in People With Minimal to Moderate Impairment Due to Chronic Stroke". en. In: *Physical Therapy* 92.6 (June 2012), pp. 791–798.

[242] C. Boulay, W. Sarnacki, J. Wolpaw, and D. McFarland. "Trained modulation of sensorimotor rhythms can affect reaction time". In: *Clinical Neurophysiology* 122.9 (2011), pp. 1820 –1826.

[243] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan. "Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface". In: *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. June 2008, pp. 2390–2397.

[244] K. K. Ang et al. "A Randomized Controlled Trial of EEG-Based Motor Imagery Brain-Computer Interface Robotic Rehabilitation for Stroke". In: *Clinical EEG and Neuroscience* 46.4 (Oct. 2015), pp. 310–320.

[245] J. W. Chung, E. Ofori, G. Misra, C. W. Hess, and D. E. Vaillancourt. "Beta-band activity and connectivity in sensorimotor and parietal cortex are important for accurate motor performance". In: *NeuroImage* 144 (Jan. 2017), pp. 164–173.

[246] N. Lebar, J. Danna, S. Moré, L. Mouchnino, and J. Blouin. "On the neural basis of sensory weighting: Alpha, beta and gamma modulations during complex movements". In: *NeuroImage* 150 (Apr. 2017), pp. 200–212.

[247] D. T. Bundy, M. Wronkiewicz, M. Sharma, D. W. Moran, M. Corbetta, and E. C. Leuthardt. "Using ipsilateral motor signals in the unaffected cerebral hemisphere as a signal platform for brain–computer interfaces in hemiplegic stroke survivors". en. In: *Journal of Neural Engineering* 9.3 (2012), p. 036011.

[248] V. Kaiser, G. Bauernfeind, A. Kreilinger, T. Kaufmann, A. Kübler, C. Neuper, and G. R. Müller-Putz. "Cortical effects of user training in a motor imagery based brain–computer interface measured by fNIRS and EEG". In: *NeuroImage*. Celebrating 20 Years of Functional Near Infrared Spectroscopy (fNIRS) 85 (Jan. 2014), pp. 432–444.

[249] E. Formaggio et al. "Modulation of event-related desynchronization in robot-assisted hand performance: brain oscillatory changes in active, passive and imagined movements". In: *Journal of NeuroEngineering and Rehabilitation* 10.1 (Feb. 2013), p. 24.

[250] C. M. Stinear, M. A. Petoe, S. Anwar, P. A. Barber, and W. D. Byblow. "Bilateral Priming Accelerates Recovery of Upper Limb Function After Stroke". EN. In: *Stroke* (Jan. 2014).

[251] C. Jeunet, B. Glize, A. McGonigal, J.-M. Batail, and J.-A. Micoulaud-Franchi. "Using EEG-based brain computer interface and neurofeedback targeting sensorimotor rhythms to improve motor skills: Theoretical background, applications and prospects". In: *Neurophysiologie Clinique* (Nov. 2018).

[252] M. A. Guadagnoli and T. D. Lee. "Challenge Point: A Framework for Conceptualizing the Effects of Various Practice Conditions in Motor Learning". In: *Journal of Motor Behavior* 36.2 (June 2004), pp. 212–224.

[253] R. VanRullen, N. Busch, J. Drewes, and J. Dubois. "Ongoing EEG phase as a trial-by-trial predictor of perceptual and attentional variability". In: *Frontiers in Psychology* 2 (2011).

[254] G. Lisi, D. Rivela, A. Takai, and J. Morimoto. "Markov Switching Model for Quick Detection of Event Related Desynchronization in EEG". English. In: *Frontiers in Neuroscience* 12 (2018).

ACKNOWLEDGMENTS

Taking a look across all the contributions of this thesis, I want to point out that this would not have been possible without the support and contributions of many people around me.

Above all, I want to express my sincere gratitude for the abundant support, the encouragement and the richness of novel ideas that I received from my Ph.D. supervisor Dr. Michael Tangermann. I highly appreciate that you introduced me to this very exciting and highly interdisciplinary research field and I thank you for giving me the opportunity to help establish your young research group as a first lab member. When I think back, your enthusiastic and optimistic attitude had a large impact on this thesis.

My grateful thanks also goes to Dr. Fabien Lotte who kindly hosted me for a research stay in his laboratory and thereby enabled this vital and enriching collaboration. Thank you for sharing many valuable methodological insights, for always providing so encouraging feedback and of course your agreement to review this thesis.

I would like to express my special appreciation to Prof. Dr. Joschka Bödecker for insightful discussions and his agreement to be an examiner of this thesis.

Moreover, I am grateful to Prof. Dr. Frank Hutter and Prof. Dr. Bernhard Nebel for their agreement on being part of my thesis committee.

I also want to warmly thank Dr. Janine Reis for sharing her clinical expertise, for highly effective meetings and for being part of our successful DFG grant writing team. A special thanks goes also to Gerd Strohmeier whose technical expertise tremendously helped to implement the SVIPT setup in our lab.

Special thanks is given to Prof. Dr. Benjamin Blankertz for providing the motor imagery dataset and valuable input for the research on SPoC regularization techniques.

I want to acknowledge the help provided by the z-office team of the BrainLinks-BrainTools Cluster of Excellence. Sincere thanks to Luise Goroncy for all her enthusiasm and the incredible energy she spent for the junior scientists activities and her support in proofreading this thesis. I also want to thank Jan Deubner for our common time as junior scientist representatives. I am glad we had the opportunity to set up all the different events, and could gain a lot of experience outside the world of research.

Moreover, I thank the support team of the NEMO cluster, especially Bernd Wiebelt and Michael Janczyk, for their straight and constant help in all different kind of issues that occasionally had to be solved.

I want to express my appreciation to all participants of the more than 100 EEG experiments, which I accompanied throughout my PhD time. I really enjoyed to welcome so many subjects in the lab for these couple of

hours. Without your patience and endurance my research would not have been possible.

I want to warmly thank my PhD colleagues David, Henrich, Sebastián and Jan for the very open atmosphere, many fruitful and intense discussions, for reading parts of this thesis, for the great times at conferences (and of course the exciting times before and after) or for having a chat when preparing a subject for the next EEG recording. I really appreciate this valuable time with you. In the same way, I want to thank Elena, Eva, Johannes, Konstantin, Oleksii, Ralvi, Stephan and Torsten for their tremendous support in evaluating novel research ideas, running a large number of experiments, for testing new implementations or simply having a coffee together.

I owe special thanks to Lena for taking the time to proof-read parts of this thesis. Thank you, Michi, for video editing the recordings with our pilot patients.

I want to express my deep thankfulness to my parents. You followed all my paths throughout my life, while always providing me the best support I can imagine and all the freedom that I needed. I also want to acknowledge my family members and friends who constantly supported me and made the time beside research so worth living. Finally, I am indebted to Mareike for her constant support and believe in me. You have always put a smile on my face.

This thesis was supported by *BrainLinks-BrainTools* Cluster of Excellence funded by the German Research Foundation (DFG, grant number EXC 1086) and the project SuitAble (DFG, grant number 387670982). The author also acknowledges support by the state of Baden-Württemberg, Germany, through bwHPC and the German Research Foundation (DFG, INST 39/963-1 FUGG).