

Über die robuste räumliche Filterung von EEG in nichtstationären Umgebungen¹

Wojciech Samek²

Abstract: Die üblichen Kommunikationsformen wie Sprache, Gestik oder Mimik erfordern eine gezielte Aktivierung der Muskulatur und können Patienten mit stark beeinträchtigter Motorik nicht (oder nur sehr eingeschränkt) zur Verfügung stehen. Gehirn-Computer-Schnittstellen ermöglichen Kommunikation durch reine Bewegungsvorstellung und stellen für diesen Personenkreis eine vielversprechende Alternative dar. Trotz einer rasanten Weiterentwicklung in den letzten Jahren, sind heutige Systeme nur sehr begrenzt für den Einsatz außerhalb einer kontrollierten Laborumgebung – also im Alltag eines Patienten – geeignet. Ein limitierender Faktor ist die mangelnde Robustheit gegenüber starkem Signalrauschen, unerwarteten Artefakten und Nichtstationarität in den Daten. Die diesem Artikel zugrunde liegende Doktorarbeit widmet sich diesem Forschungsfeld und stellt neue Entwicklungen bei der räumlichen Filterung von Hirnsignalen, einem zentralen Schritt bei der Informationsverarbeitung in Gehirn-Computer-Schnittstellen, vor. Die Beiträge der Arbeit reichen von der Entwicklung neuartiger Regularisierungsstrategien für die räumliche Filterung bei Nichtstationarität über die Herleitung speziell auf die Daten abgestimmter robuster Parameterschätzverfahren bis hin zur Formulierung einer allgemeinen informationsgeometrischen Sichtweise auf die räumliche Filterung. Die vorgestellten Konzepte und Methoden finden weit über das Feld der Gehirn-Computer-Schnittstellen hinaus Anwendung und können zur Lösung verschiedener Probleme in der Informatik und Mathematik herangezogen werden.

1 Gehirn-Computer-Schnittstellen

Kommunikation ist sowohl in der Informatik als auch für uns Menschen von enormer Bedeutung. Während die Informationsweitergabe im Computer durch das Verschieben von Bits und Bytes erfolgt, sind Menschen üblicherweise auf Kommunikationsformen wie Sprache, Gestik und Mimik angewiesen. Die Fähigkeit bestimmte Muskelgruppen gezielt aktivieren zu können, ist Grundvoraussetzung für diese Arten der menschlichen Kommunikation. Krankheiten wie die amyotrophe Lateralsklerose (ALS) können die Motorik eines Menschen so stark beeinträchtigen, dass er die Fähigkeit zur gezielten Muskelsteuerung und zur Kommunikation vollständig verliert. Ein solcher Zustand wird *locked-in* genannt.

Gehirn-Computer-Schnittstellen (*engl.* Brain-Computer Interfaces oder BCIs) ermöglichen eine neue Art der Kommunikation, die weder Muskelaktivität noch ein funktionierendes zentrales Nervensystem voraussetzt und somit auch schwer kranken ALS-Patienten zur Verfügung steht. Bereits vor über 40 Jahren kam die Idee auf, dass Hirnaktivität mittels Enzephalographie (EEG) aufgenommen und zur Kommunikation mit der Umgebung oder zur Kontrolle technischer Geräte verwendet werden könnte [Vi73]. Die ersten Systeme

¹ Englischer Titel der Dissertation: “On robust spatial filtering of EEG in nonstationary environments”

² Fraunhofer Heinrich-Hertz-Institut, AG Maschinelles Lernen, Einsteinufer 37, 10587 Berlin, wojciech.samek@hhi.fraunhofer.de

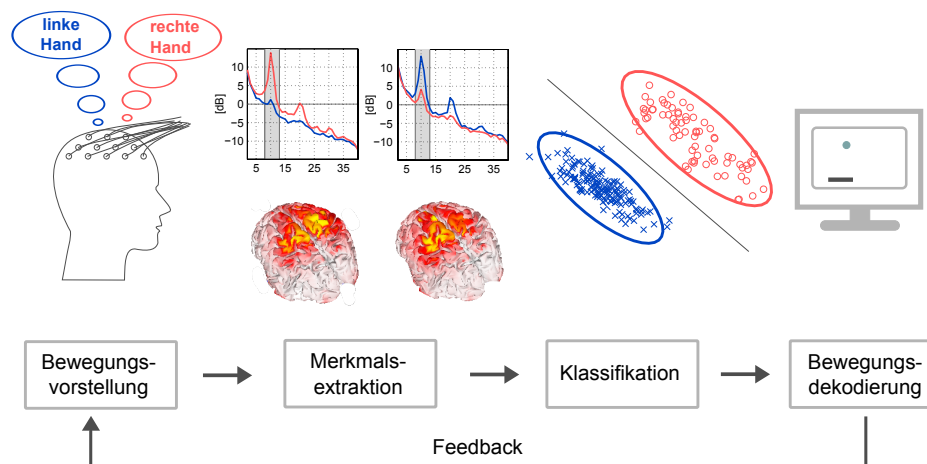


Abb. 1: Arbeitsweise eine BCIs. Der Proband stellt sich eine Bewegung vor und diese induziert eine Veränderung in der Stärke von Oszillationen in bestimmten Frequenzbereichen und Hirnregionen. Probandenspezifische Merkmale, welche diesen Effekt gut abbilden, werden aus dem gemessenen EEG-Signal extrahiert und anschließend zur Dekodierung der vorgestellten Bewegungen verwendet. Das Ergebnis der Dekodierung wird schließlich dem Benutzer als Feedback zurückgegeben.

dieser Art basierten auf dem Neurofeedback-Prinzip und erforderten wochenlanges Training. Der Einsatz von Verfahren des Maschinellen Lernens [MAB03] revolutionierte das Forschungsfeld und machte BCI-Systeme erstmals praktikabel, denn diese benötigten fortan lediglich eine Kalibrierungszeit von wenigen Minuten, um sich optimal auf die einzelnen Benutzer einzustellen. Heute verwenden fast alle BCI-Systeme diese datengetriebenen Techniken.

Im Folgenden wird die Arbeitsweise eines BCI-Systems vorgestellt. Siehe dazu auch Abbildung 1. Wie bereits erwähnt, gibt es bei den hier beschriebenen Gehirn-Computer-Schnittstellen eine Kalibrierungsphase, in der die probandenspezifischen Parameter gelernt werden, und eine Testphase, in der das System angewendet wird. In der Kalibrierungsphase erhält der Proband abwechselnd den Auftrag sich eine Bewegung mit der linken oder rechten Hand bzw. mit den Füßen vorzustellen. Aus der Neurophysiologie ist bekannt, dass die bloße Vorstellung einer Bewegung bereits zu Veränderungen in der neuronalen Aktivität in bestimmten Bereichen des sensomotorischen Cortex führt, welche im EEG als Änderungen in der Stärke von Oszillationen in den Frequenzbereichen 8-13 Hz (μ -Band) und 18-30 Hz (β -Band) sichtbar werden. Typischerweise findet eine Desynchronisation kontralateral zur vorgestellten Bewegung und eine ipsilaterale Synchronisation statt. Dies ist bekannt als der Event-Related Desynchronization/Synchronization (ERD/ERS) Effekt [PLdS99]. Eine zentrale Aufgabe der verwendeten Algorithmen ist die Extraktion von probandenspezifischen Merkmalen, welche die mit der Bewegungsvorstellung einhergehende Aktivität möglichst gut abbilden. Anschließend folgt die Klassifikation dieser Merkmale, d.h. die Dekodierung der vorgestellten Bewegung. Auch in diesem Schritt müssen die optimalen Parameter zunächst mit Hilfe der Kalibrierungsdaten gelernt

werden. In der Testphase wird das BCI-System mit den gelernten Parametern angewendet und das Ergebnis der Dekodierung wird dem Benutzer als Feedback (z.B. durch die Bewegung eines Punktes auf dem Bildschirm) zurückgegeben.

Ein großes Problem von Gehirn-Computer-Schnittstellen ist die fehlende Robustheit gegenüber starkem Rauschen, unerwarteten Artefakten und zeitabhängigen Veränderungen der Verteilung der Daten (*sog.* Nichtstationarität). Nur ein sehr kleiner Teil der im EEG abgebildeten neuronalen Aktivität ist BCI-spezifisch, der überwiegende Teil entstammt ganz anderen parallel laufenden neuronalen Prozessen, die beispielsweise für die Steuerung von Körperfunktionen oder die Verarbeitung von Sinneseindrücken verantwortlich sind. Diese nicht BCI-spezifische Aktivität kann das für die Kommunikation relevante Signal deutlich überlagern und kleinste Änderungen können zu Nichtstationarität führen. Zusätzlich können Faktoren wie eine fortschreitende Müdigkeit des Benutzers, Aufmerksamkeitsschwankungen, Strategiewechsel bei der Durchführung der vorgestellten Bewegung, Änderungen in der Impedanz sowie veränderter sensorischer Input das gemessene Signal sehr stark beeinflussen und die statistischen Eigenschaften der Daten verändern. Bewegungen aller Art (Muskel- und Augenbewegungen, Blinzeln, Schlucken etc.) sind ebenfalls ein großes Problem. Diese Artefakte beeinflussen die datengetriebenen Methoden des Maschinellen Lernens und haben oft negative Auswirkungen auf die Performanz. In nicht kontrollierten Umgebungen (z.B. zu Hause, auf der Straße) gibt es noch viel mehr Ablenkungs- und Fehlerquellen, welche die Durchführung eines BCI-Experimentes erheblich erschweren. Die Entwicklung von robusten Techniken der EEG-Analyse ist somit ausschlaggebend für den zukünftigen Erfolg von Gehirn-Computer-Schnittstellen.

2 Räumliche Filterung

Die räumliche Filterung ist ein zentraler Bestandteil der Informationsverarbeitung in einem BCI-System. Sie verbessert das Signal-Rausch-Verhältnis und macht oftmals die von der Bewegungsvorstellung induzierte neuronale Aktivität erst sichtbar. Im Folgenden wird der CSP Algorithmus – eine sehr populäre Methode zur Berechnung von räumlichen Filtern – sowie zwei neue Ansätze zur Filterung bei Nichtstationarität vorgestellt.

2.1 Der CSP Algorithmus

Der Common Spatial Patterns (CSP) Algorithmus [BI08] zielt auf die Extraktion von Signalquellen mit dem stärksten ERD/ERS-Effekt ab. Dazu findet CSP räumliche Filter, die den Quotienten aus den Varianzen der beiden Klassen (z.B. Bewegungsvorstellung mit linken und rechten Hand) maximieren. Mathematisch entspricht dies der Lösung eines generalisierten Eigenwertproblems. Da CSP zur Überanpassung neigt und gegenüber Artefakten nicht robust ist, wurden verschiedene Erweiterungen des Algorithmus präsentiert. Eine wichtige Kategorie bilden dabei die regularisierten CSP Methoden [LG11]. Diese Algorithmen lassen sich effizient berechnen und maximieren die folgende Zielfunktion

$$w^* = \operatorname{argmax}_w \frac{w^\top \Sigma_c w}{w^\top (\Sigma_1 + \Sigma_2 + \lambda K) w}, \quad (1)$$

wobei $\Sigma_c \in \mathbb{R}^{D \times D}$ die Kovarianzmatrix der Klasse $c = \{1, 2\}$, $\lambda \geq 0$ der Regularisierungsparameter und K die Regularisierungsmatrix sind. Durch verstärkte Regularisierung werden die berechneten räumlichen Filter w^* mehr und mehr orthogonal zur Matrix K (im Extremfall gilt $w^{*\top} K w^* = 0$). Bei entsprechender Wahl von K können so verschiedene Eigenschaften wie z.B. eine Reduzierung der Norm (K ist Einheitsmatrix) forciert werden.

2.2 Regularisierung bei Nichtstationarität

Der vorgestellte CSP Algorithmus maximiert zwar den Quotienten aus den beiden (durchschnittlichen) Varianzen, aber er ignoriert die Variabilität der extrahierten Merkmale innerhalb der einzelnen Klassen. Dies kann zu einer nichtstationären Merkmalsverteilung führen und die nachfolgende Klassifizierung deutlich erschweren. Ein einfaches Maß für die Variabilität der Merkmale ist die mittlere absolute Abweichung vom Mittelwert

$$P(w) = \frac{1}{2n} \sum_{c=1}^2 \sum_{i=1}^n \left| w^\top \Sigma_c^i w - w^\top \Sigma_c w \right|, \quad (2)$$

wobei die Kovarianzmatrix vom i -ten Versuchstrial der Klasse c mit Σ_c^i und die Anzahl der Versuchstrials pro Klasse mit n dargestellt werden. Diese Regularisierungsfunktion kann nicht mit der regularisierten CSP Methode kombiniert werden, weil sich $P(w)$ nicht als $w^\top K w$ darstellen lässt. Folgende Approximation macht jedoch eine solche Darstellung möglich und ist Grundlage des stationary CSP (sCSP) Algorithmus [Sa12]

$$\left| w^\top \Sigma_c^i w - w^\top \Sigma_c w \right| \approx w^\top \mathcal{F}(\Sigma_c^i - \Sigma_c) w, \quad (3)$$

wobei \mathcal{F} ein Operator ist, der symmetrischen Matrizen durch die Umkehrung der Vorzeichen negativer Eigenwerte positive Definitheit verleiht. Es stellt sich nun die Frage, ob die Approximation von $P(w)$ die Nichtstationarität der extrahierten Merkmale (gemessen als mittlere absolute Abweichung) in gleichem Maße reduziert wie $P(w)$. Unter gewissen Annahmen ist dies tatsächlich der Fall.

Theorem 1 Wenn die Kovarianzmatrizen $\Sigma_c^i = V^\top D_{i,c} V$ sich als Produkt einer allgemeinen orthogonalen Matrix V und einer spezifischen Diagonalmatrix $D_{i,c}$ darstellen lassen, d.h. gemeinsam diagonalisierbar sind, dann gilt

$$\operatorname{argmax}_w \sum_{c=1}^2 \sum_{i=1}^n \left| w^\top \Sigma_c^i w - w^\top \Sigma_c w \right| = \operatorname{argmax}_w w^\top \underbrace{\left(\sum_{c=1}^2 \sum_{i=1}^n \mathcal{F}(\Sigma_c^i - \Sigma_c) \right)}_K w$$

Die zweite Regularisierungsstrategie – stationary subspace CSP (ssCSP) – hat die Reduzierung der Veränderung der Merkmalsverteilung zwischen der Kalibrierungs- und Testphase zum Ziel. Diese Art von Nichtstationarität tritt häufig auf und wird vielfach als Grund für schlechte Performanz genannt [Sh06]. Die Regularisierung beruht auf der Beobachtung, dass Änderungen in der Merkmalsverteilung zwischen der Kalibrierungs- und

Testphase probandenunspecific sind [SMM13]. Eine mögliche Erklärung für dieses Phänomen ist eine bei allen Probanden auftretende Änderung im experimentellen Ablauf. Das könnte beispielsweise die Verwendung von visuellen Feedback in der Testphase sein, welcher in der Kalibrierungsphase fehlte und bei allen Probanden zu ähnlichen Aktivierungsmustern und somit zu ähnlicher Nichtstationarität führt. Um die räumliche Filterung gegenüber solcher Nichtstationarität invariant zu machen, muss eine passende Matrix K konstruiert werden. Hierfür werden zunächst individuelle Filter extrahiert, die starke Änderungen in der Merkmalsverteilung hervorrufen. Anschließend wird mit einem auf der Singulärwertzerlegung basierenden Ansatz ein probandenunspecificer Nichtstationaritätsraum konstruiert, der sich schließlich mit Matrix K darstellen lässt.

3 Informationsgeometrische Methoden

Nachdem zwei neue auf Regularisierung basierende Methoden der räumlichen Filterung vorgestellt wurden, wird im Folgenden das Problem der algorithmischen Robustheit aus einer allgemeineren Perspektive betrachtet. Zudem wird in diesem Kapitel eine Vielzahl von grundlegenden neuer Verfahren der räumlichen Filterung präsentiert, die alle auf sog. *Divergenzen* basieren. Divergenzen sind wichtige Elemente aus der Informationsgeometrie [AN00], die als Maß für die Diskrepanz zwischen zwei Wahrscheinlichkeitsverteilungen aufgefasst werden können. Die vorgestellten Konzepte und Methoden sind nicht auf räumliche Filterung im BCI beschränkt, sondern können auch zur Lösung verschiedener Probleme in der Informatik und Mathematik herangezogen werden.

3.1 Parameterschätzung mit Divergenzen

Das Schätzen von Parametern ist ein wichtiger Schritt in vielen Anwendungen und Algorithmen, inklusive der Berechnung von räumlichen Filtern. Eine wichtige Klasse von Parameterschätzern sind die sogenannten Maximum-Likelihood (ML) Methoden. Man kann leicht zeigen [EK01], dass die Maximierung der Likelihood-Funktion und die Minimierung der Kullback-Leibler (KL) Divergenz zwischen der empirischen Verteilung p und einem Modell q_θ zur gleichen Schätzung $\hat{\theta}$ führt. Dieser Parameterschätzer ist nicht robust, weil ein einzelner Datenpunkt die Schätzung signifikant verfälschen kann. Aus der Literatur [EK01] ist bekannt, dass die Verwendung der Beta Divergenz³ – einer Verallgemeinerung der KL Divergenz – den Einfluss von Ausreißern wesentlich reduziert (d.h. diese heruntergewichtet) und somit zu robusten Schätzungen führt. Die Verwendung dieser Divergenz mit einem Gauss-Modell $q \sim \mathcal{N}(0, \Sigma)$ führt zu einer robusten Schätzung der Kovarianzmatrizen [EK01], welche anschließend im CSP Algorithmus verwendet werden können. Bei dieser Art der robusten Filterung werden einzelne Messpunkte des EEG-Signals heruntergewichtet und als Ausreißer behandelt. Für BCI-Daten ist jedoch noch eine weitere Art der Robustheit denkbar, nämlich die Robustheit gegenüber ganzen Versuchstrials, die Ausreißer darstellen [SK14]. Diese neue Art der Robustheit lässt sich unter

³ Die Beta Divergenz ist definiert als $D_\beta(p || q) = \frac{1}{\beta} \int (p^\beta - q^\beta) dx - \frac{1}{\beta+1} \int (p^{\beta+1} - q^{\beta+1}) dx$

Verwendung der Wishart-Verteilung $q \sim \mathcal{W}(\nu, \Sigma)$ herleiten. Die Beta Divergenz zwischen der empirischen Verteilung der versuchsweisen Streumatrizen $S_i = \sum_{t=1}^T x_i(t)x_i(t)^\top$ und dem Wishart-Model mit Parametern ν und Σ lässt sich wie folgt iterativ minimieren

$$\Sigma^{(k+1)} = \frac{\sum_{i=1}^n \psi_\beta \left(\ell \left(S_i; \Sigma^{(k)}, \nu \right) \right) S_i}{\nu \sum_{i=1}^n \psi_\beta \left(\ell \left(S_i; \Sigma^{(k)}, \nu \right) \right) - \gamma |\Sigma^{(k)}|^{\frac{(\nu-D-1)\beta}{2}}} \quad (4)$$

mit $\psi_\beta \left(\ell(S; \Sigma, \nu) \right) = |S|^{\frac{(\nu-D-1)\beta}{2}} e^{-\text{tr}(\frac{\beta}{2}\Sigma^{-1}S)}$ als Faktor, der Ausreißer-Versuchstrials (parametrisiert durch Streumatrix S) heruntergewichtet, und einer Konstante γ . Für $\beta \rightarrow 0$ und $\nu = T$ erhält man die gewöhnliche (nicht robuste) ML-Schätzung der Kovarianzmatrix $\Sigma = \frac{1}{n} \sum_{i=1}^n \Sigma^i = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T x_i(t)x_i(t)^\top$. Für Parameter $\beta > 0$ führt der neue Kovarianzmatrixschätzer in Kombination mit CSP zu einer robusten (in Hinblick auf Ausreißer-Versuchstrials) Berechnung von räumlichen Filtern.

3.2 Räumliche Filterung mit Divergenzen

Ein wichtiger Beitrag der zugrunde liegenden Dissertation [Sa14] ist die Formulierung des CSP Algorithmus als Divergenzmaximierungsproblem [Sa13, SKM14]. Ein großer Vorteil dieser Betrachtungsweise ist, dass man durch die Ausnutzung der speziellen Eigenschaften (z.B. Robustheit) anderer Divergenzen oder durch die Verwendung anderer Wahrscheinlichkeitsmodelle (z.B. Studentische t-Verteilung) ganz neue Varianten des ursprünglichen Algorithmus erhält. Ferner bietet die Art der verwendeten Optimierung (Berechnung der Lösung in einem Schritt vs. schrittweise Berechnung der Lösung) eine Flexibilität, die bei der gewöhnlichen Formulierung eines Algorithmus oftmals nicht vorhanden ist. Dieser *Divergenztrick* [SKM14] wird hier zwar für CSP angewendet, er kann aber ohne Einschränkung für viele andere bekannten Algorithmen wie PCA, SSA oder CCA ebenfalls benutzt werden, um neuartige Varianten dieser Methoden herzuleiten, mit denen man Probleme in der Informatik u. U. besser lösen kann. Das folgende Theorem ist die theoretische Basis aller in diesem Abschnitt beschriebener Methoden.

Theorem 2 Wenn $W \in \mathbb{R}^{D \times d}$ die d diskriminativsten mit CSP berechneten räumlichen Filter sind und V sich als $V^\top = I_d R P \in \mathbb{R}^{d \times D}$ darstellen lässt mit $P(\Sigma_1 + \Sigma_2)P^\top = I$, $R^\top R = I \in \mathbb{R}^{D \times D}$ und einer abgeschnittenen Identitätsmatrix $I_d \in \mathbb{R}^{d \times d}$, dann gilt

$$\text{span}(W) = \text{span}(V^*) \quad (5)$$

$$\text{mit } V^* = \text{argmax}_V \tilde{D}_{kl} \left(\mathcal{N}(0, V^\top \Sigma_1 V) \parallel \mathcal{N}(0, V^\top \Sigma_2 V) \right), \quad (6)$$

wobei \tilde{D}_{kl} für die symmetrische KL Divergenz steht und $\mathcal{N}(\mu, \Sigma)$ eine Gauss-Verteilung mit Mittelwert μ und Kovarianzmatrix Σ repräsentiert. Das Theorem zeigt also, dass die Berechnung von räumlichen Filtern auch als informationsgeometrisches Optimierungsproblem aufgefasst werden kann. Um Robustheit gegenüber Artefakten zu forcieren, kann man entweder den zuvor vorgestellten robusten Parameterschätzer verwenden oder die

Zielfunktion (6) entsprechend anpassen und die symmetrische KL Divergenz durch eine robuste Divergenz ersetzen. Die Verwendung der symmetrischen Beta Divergenz führt dazu, dass Ausreißer-Terme heruntergewichtet werden, d.h. weniger Einfluss auf das Ergebnis haben.

Die divergenzbasierte Formulierung erlaubt auch die Kombination der Zielfunktion mit einem Regularisierungsterm. Dadurch können spezielle Eigenschaften wie Stationarität forciert werden. Ferner lassen sich bequem Daten von anderen Versuchspersonen in die Optimierung mit einbeziehen, indem diese als zusätzliche Divergenzterme beschrieben werden. Mit der Verwendung des Parameters β lassen sich Ausreißer-Effekte von einzelnen Versuchspersonen oder einzelne Versuchstrials entweder hervorheben oder heruntergewichten. Wie zuvor erwähnt, lassen sich auch ganze Unterräume (die ganze Matrix V) in einem Schritt berechnen. Diese Flexibilität führt zu vielen neuen, für bestimmte Anwendungsfälle optimalen Verfahren der räumlichen Filterung und wurde in [SKM14] ausführlich diskutiert und evaluiert.

4 Experimentelle Evaluation

Zur Evaluation der vorgestellten Methoden wird der Vital-BCI Datensatz [B110] mit 80 Versuchspersonen verwendet. Es werden 62 EEG-Kanäle ausgewählt und die Daten werden im Frequenzbereich 8-30 Hz spektral gefiltert. Die Merkmalsextraktion basiert auf der neuronalen Aktivität im Zeitintervall 750-3500ms nach dem Stimulus. Sechs räumliche Filter und ein linearer Klassifikator (LDA) werden zum Dekodieren der vorgestellten Bewegung benutzt. Die methodenspezifischen Parameter werden mit Hilfe von Kreuzvalidierung gewählt. Der Vital-BCI Datensatz ist aus mindestens drei Gründen für die Evaluation der vorgestellten robusten Methoden besonders gut geeignet. (1) Die große Anzahl von Versuchspersonen ermöglicht einen statistisch validen Vergleich der Performanz der Methoden. (2) Die Teilnehmer der Vital-BCI Studie haben zuvor noch nie ein BCI bedient. Die Daten sind dadurch i.A. deutlich verrauschter (Unsicherheiten der Probanden, Bewegungsartefakte, Nichtstationarität durch Wechsel der Strategie, Lerneffekte) als die eines erfahrenen BCI-Benutzers. Für die Einschätzung der BCI-Performanz in realer Umgebung ist ein solcher Datensatz ideal geeignet. (3) Der Datensatz wurde bereits in einer Vielzahl von früheren Studien verwendet, so dass Performanzwerte für verschiedene Methoden vorhanden sind. Dies erhöht die Vergleichbarkeit der hier präsentierten Ergebnisse.

Im Folgenden wird eine Auswahl der Ergebnisse der Dissertation [Sa14] präsentiert. Bei Verwendung von CSP beträgt der durchschnittliche Dekodierungsfehler für die 80 Probanden 31.3%. Die Bandbreite der Fehlerraten ist dabei sehr groß und reicht von 2.5% bis 52.3%. Nur etwa die Hälfte der Probanden ist in der Lage das BCI-System in einer Weise zu kontrollieren, die effiziente Kommunikation ermöglicht. Der Grenzwert hierfür liegt bei einer Dekodierungsfehlerrate von 30%. Durch die Verwendung der in diesem Artikel vorgestellten robusten Methoden lässt sich eine deutliche Senkung der Fehlerrate für die meisten Probanden erreichen. Für sCSP beträgt der durchschnittliche Performanzgewinn fast 10%. Für einzelne Probanden fällt die Verbesserung deutlich größer aus, z.B. sinkt der Dekodierungsfehler bei den Probanden *VPtbo* und *VPkl* von 48.6% auf 22.0% bzw.

von 40.0% auf 17.7%. Die Anwendung von sCSP versetzt nicht nur die beiden Benutzer erst in die Lage ein BCI effizient (d.h. mit Fehlerrate kleiner 30%) zu bedienen, sondern ermöglicht insgesamt 10 Benutzern den Sprung über diesen wichtigen Grenzwert. Der einseitige Wilcoxon-Vorzeichentest zeigt zudem, dass der Performanzgewinn mit einem p-Wert kleiner als 10^{-4} höchst signifikant ist. Die Nichtstationarität der Merkmalsverteilung nimmt bei Anwendung von sCSP für die meisten Probanden (verglichen mit CSP) ab. Für den Parameter $\lambda = 0.25$ beträgt diese Abnahme durchschnittlich ca. 50%.

Die andere hier vorgestellte Regularisierungsstrategie verbessert die Performanz, indem sie die Nichtstationarität zwischen Kalibrierungs- und Testphase reduziert. Abbildung 2 (a) veranschaulicht den Effekt dieser Regularisierung. Jeder Punkt repräsentiert einen Versuchstrial. Für CSP sieht man eine deutliche zeitliche Abhängigkeit der Merkmalsverteilung, die zu einer großen Diskrepanz zwischen der Verteilungen der Kalibrierungs- (leere Punkte) und Test-Merkmale (volle Punkte) und somit zu sehr schlechter Klassifizierbarkeit führt. Der ssCSP Algorithmus benutzt Daten von anderen Versuchspersonen, um die auftretenden Veränderungen zu antizipieren und invariante räumliche Filter zu berechnen. Im dargestellten Fall wird die Diskrepanz zwischen der Verteilung der Kalibrierungs- und Testdaten durch die Regularisierung deutlich reduziert. Dieses Ergebnis zeigt, dass ein Transfer von Informationen über Nichtstationaritäten möglich ist und zu einer Senkung des Dekodierungsfehlers (z.B. für *VPjs* von 34% auf 24%) führt.

Bisher wurde gezeigt, dass die Reduzierung der Nichtstationarität zu besseren Dekodierungsergebnissen führt. Im Folgenden steht die erhöhte Robustheit gegenüber Artefakten im Fokus der Evaluierung. Der im Abschnitt 3.1 vorgestellten robuste Parameterschätzer führt – wenn mit CSP kombiniert – zu einer signifikanten ($p = 0.0046$) Senkung der Fehlerrate. Einen noch stärkeren Effekt ($p < 10^{-4}$) erhält man bei Verwendung des auf Beta Divergenz basierten Algorithmus β -divCSP. Auch hier erzielen einzelne Probanden enorme Verbesserungen in der Performanz, z.B. sinkt die Fehlerrate bei dem Probanden *VPtbo* oder *VPlib* von 48.6% auf 11.0% bzw. von 32.0% auf 16.7%. Die durch erhöhte Stationarität und Robustheit erzielten Verbesserungen sind i.A. nicht gleich, im Gegenteil beide Ansätze sind zu einem gewissen Grad komplementär zueinander. Bei manchen Versuchspersonen führt eine erhöhte Robustheit zu einer deutlichen Senkung der Fehlerrate, bei anderen Probanden hingegen ist Nichtstationarität das primäre Problem. Abbildung 2 (b) veranschaulicht den Einfluss von Artefakten auf die Filterberechnung. Anstatt die durch Bewegungsvorstellung induzierte neuronale Aktivität zu extrahieren, konzentriert sich CSP – aufgrund eines Ausreißer-Versuchstrials – auf das Signal der FC5 Elektrode. Die robuste β -divCSP Methode schenkt einzelnen Ausreißereffekte weniger Beachtung und ist in der Lage, die von der Bewegungsvorstellung induzierte Aktivität zu extrahieren.

In der Dissertation werden die vorgeschlagenen Algorithmen mit weiteren Referenzmethoden verglichen. Auch diese neueren Methoden der räumlichen Filterung sind anfällig für Artefakte und Nichtstationarität und führen zu signifikant größeren Fehlerraten als die hier vorgestellten Algorithmen, welche aus Nicht-BCI-Benutzern (Fehlerrate über 30%) sichere Verwender dieser Zukunftstechnologie machen können.

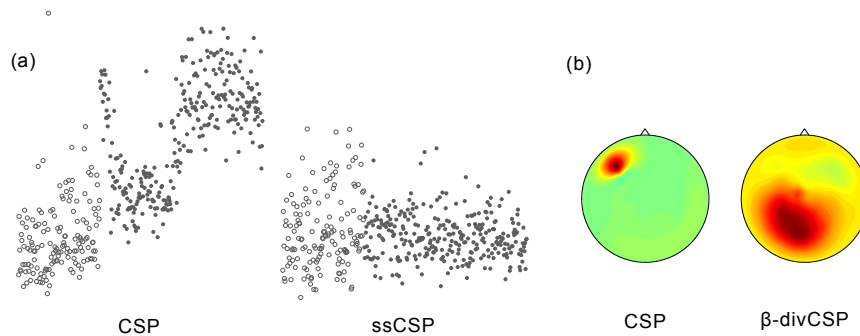


Abb. 2: (a) Jeder Punkt repräsentiert einen Kalibrierungs- (leerer Punkt) oder Test-Versuchstrials (voller Punkt). ssCSP reduziert die Nichtstationarität zwischen beiden Verteilungen deutlich. (b) Ein Ausreißer im FC5 Kanal führt dazu, dass CSP die relevante Aktivität nicht extrahiert. Die robuste Methode hingegen reduziert den Einfluss des Ausreißers und zeigt das erwartete Aktivierungsmuster.

5 Zusammenfassung und Ausblick

Die in diesem Artikel vorgestellten Methoden und Konzepte zur robusten räumlichen Filterung können die Performanz von BCI-Systemen deutlich steigern und werden durch den Trend, Gehirn-Computer-Schnittstellen auch in nicht kontrollierten, realistischen Umgebungen einzusetzen, zukünftig eine zentrale Rolle in der BCI-Forschung einnehmen. Diese Methoden erweitern die dem Stand der Technik entsprechenden Algorithmen nicht nur dahingehend, dass sie Stationarität forcieren und die Robustheit gegenüber Datenpunkt- und Versuchstrial-Artefakten signifikant erhöhen, sondern sie erlauben auch eine neue Art des Informationstransfers zwischen Versuchspersonen, nämlich den Austausch über die erwarteten Nichtstationaritäten in den Daten. Ähnlich wie der Kernel-Trick [MAB03] die Informatik revolutioniert, weil er aus linearen Algorithmen nicht-lineare Methoden macht, ist zu hoffen, dass der hier eingeführte Divergenz-Trick, der durch die Verwendung von bestimmten Divergenzen und Wahrscheinlichkeitsmodellen aus nicht-robusten Methoden robuste Algorithmen macht oder andere wünschenswerte Eigenschaften (z.B. Stationarität) forciert, zukünftig auch einen ähnlich großen Einfluss haben wird.

Literaturverzeichnis

- [AN00] Amari, S.; Nagaoka, H.: Methods of information geometry. Jgg. 191 in Translations of Mathematical Monographs. Oxford Univ. Press, 2000.
- [BI08] Blankertz, B.; Tomioka, R.; Lemm, S.; Kawanabe, M.; Müller, K.-R.: Optimizing Spatial filters for Robust EEG Single-Trial Analysis. IEEE Sig. Proc. Mag., 25(1):41–56, 2008.
- [BI10] Blankertz, B.; Sannelli, C.; Halder, S.; Hammer, E. M.; Kübler, A.; Müller, K.-R.; Curio, G.; Dickhaus, T.: Neurophysiological Predictor of SMR-Based BCI Performance. NeuroImage, 51(4):1303–1309, 2010.

- [EK01] Eguchi, S.; Kano, Y.: Robustifying maximum likelihood estimation. Tokyo Institute of Statistical Mathematics, Tech. Rep, 2001.
- [LG11] Lotte, F.; Guan, C.: Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms. *IEEE Trans. Biomed. Eng.*, 58(2):355–362, 2011.
- [MAB03] Müller, K.-R.; Anderson, C. W.; Birch, G. E.: Linear and Non-Linear Methods for Brain-Computer Interfaces. *IEEE Trans. Neur. Sys. & Reh. Eng.*, 11(2):165–169, 2003.
- [PLdS99] Pfurtscheller, G.; Lopes da Silva, F. H.: Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophys.*, 110(11):1842–1857, 1999.
- [Sa12] Samek, W.; Vidaurre, C.; Müller, K.-R.; Kawanabe, M.: Stationary Common Spatial Patterns for Brain-Computer Interfacing. *J. Neural Eng.*, 9(2):026013, 2012.
- [Sa13] Samek, W.; Blythe, D.; Müller, K.-R.; Kawanabe, M.: Robust spatial filtering with beta divergence. In: *Adv. in NIPS*. S. 1007–1015, 2013.
- [Sa14] Samek, W.: On robust spatial filtering of EEG in nonstationary environments. Dissertation, Technische Universität Berlin, 2014.
- [Sh06] Shenoy, P.; Krauledat, M.; Blankertz, B.; Rao, R. P. N.; Müller, K.-R.: Towards Adaptive Classification for BCI. *J. Neural Eng.*, 3(1):R13–R23, 2006.
- [SK14] Samek, W.; Kawanabe, M.: Robust Common Spatial Patterns by Minimum Divergence Covariance Estimator. In: *Proc. of ICASSP*. S. 2059–2062, 2014.
- [SKM14] Samek, W.; Kawanabe, M.; Müller, K.-R.: Divergence-based Framework for Common Spatial Patterns Algorithms. *IEEE Rev. Biomed. Eng.*, 7:50–72, 2014.
- [SMM13] Samek, W.; Meinecke, F. C.; Müller, K.-R.: Transferring Subspaces Between Subjects in Brain-Computer Interfacing. *IEEE Trans. Biomed. Eng.*, 60(8):2289–2298, 2013.
- [Vi73] Vidal, J. J.: Toward Direct Brain-Computer Communication. *Annual Rev. of Biophys. and Bioeng.*, 2(1):157–180, 1973.



Wojciech Samek hat an der Humboldt-Universität zu Berlin Informatik studiert und war von 2010-2014 Doktorand am Lehrstuhl Maschinelles Lernen der Technischen Universität Berlin. Nachdem er seine Dissertation mit Auszeichnung abgeschlossen hat, wurde er im Oktober 2014 mit dem Aufbau und der Leitung der neuen Arbeitsgruppe Machine Learning am Fraunhofer Heinrich-Hertz-Institut betraut. Wojciech Samek war Stipendiat der Studienstiftung des deutschen Volkes und des Graduiertenkollegs GRK 1589 sowie Ph.D. Fellow am BCCN Berlin. Zu seinen Auslandserfahrungen zählen ein Auslandssemester an der Heriot-Watt University und der University of Edinburgh (2007-2008), ein Forschungsaufenthalt am NASA Ames Research Center in Mountain View (2009) sowie mehrere Kurzaufenthalte bei ATR International in Kyoto (2012 und 2013). Wojciech Samek ist Autor von mehr als 30 Publikationen und ist als Reviewer für Konferenzen und Journale (z.B. NIPS oder Proceedings of the IEEE) tätig. Seine Forschungsinteressen liegen im Bereich Maschinelles Lernen, Gehirn-Maschinen-Schnittstellen und Maschinelles Sehen.