

EEG-Workload-Klassifikation mit 8-Kanal-Mobil-EEG – Pipeline und Empfehlungen

Aktueller Stand und Best Practices

Die **Klassifikation mentaler Belastung (Workload) mit EEG** ist ein etabliertes Forschungsthema in der Brain-Computer-Interface und Neuroergonomie. Bereits Brouwer et al. (2012) zeigten in einem n-Back-Arbeitsgedächtnistest, dass **EEG-Spektralmaße (v.a. Theta- und Alpha-Bandleistung) sowie ERP-Komponenten (P300)** zuverlässig zwischen hoher und niedriger mentaler Last unterscheiden können ¹. In ihrer Studie erreichten Klassifikationsmodelle für die höchste vs. niedrigste Belastungsstufe **Genauigkeiten von 80–90 %** ¹ – ein frühes Indiz dafür, dass EEG-Signale die kognitive Belastung widerspiegeln. Ein gut reproduzierter Befund ist, dass bei steigender mentaler Belastung **die Theta-Leistung (insbesondere frontal) zunimmt und die Alpha-Leistung (insbesondere parietal/okzipital) abnimmt** ². Diese Kombination – *erhöhte frontale Theta-Aktivität bei gleichzeitig unterdrückter Alpha-Rhythmik* – wird konsequent als EEG-Korrelat von Workload berichtet (vgl. Brouwer et al., 2012 ²). Ebenso wurde häufig eine Zunahme schnellerer Bänder beobachtet (Beta-/Gamma-Power steigen mit Aufgabenlast) ³, während Alpha als dominantes Grundrhythmusband unter kognitiver Last reduziert ist ⁴.

Neuere Arbeiten bestätigen diese Trends auch mit tragbaren EEG-Systemen. So erzielte Wang et al. (2016) mit einem 14-Kanal-Wireless-EEG (z. B. Emotiv EPOC) etwa **81 % Genauigkeit** in der Workload-Erkennung ⁵. Matthews et al. (2017) erreichten ~80 % mit nur 9 Elektroden ⁵. Diese Ergebnisse zeigen, dass auch mobile/weniger-kanalige EEG-Geräte nützliche Workload-Information liefern können. Allerdings ist bekannt, dass die **Generalisierung über Personen und Sitzungen hinweg** eine große Herausforderung darstellt: Training und Test auf unterschiedlichen Tagen oder Probanden verschlechtern typischerweise die Performance deutlich ⁶. Beispielsweise berichteten Adewale und Panoutsos (2021) mit einem 14-Kanal-Headset *innerhalb* einer Session eine Klassifikationsgenauigkeit von ~98,5 % (für niedrige vs. hohe Last), jedoch nur ~74,4 % *über Probanden hinweg* ⁷. Dieses Gefälle unterstreicht die Notwendigkeit guter Praktiken für robuste, personenübergreifende Modelle.

Best Practices: Die vorliegende Pipeline enthält bereits wichtige Vorverarbeitungsschritte, die dem Stand der Technik entsprechen. Ein **Notch-Filter** zur Unterdrückung von Netzbrumm (50 Hz) und ein **Bandpass-Filter** (typisch ca. 1–30 Hz bzw. 1–40 Hz) entfernen hoch- und niederfrequentes Rauschen außerhalb der relevanten EEG-Bänder. Anschließend sorgt die **unabhängige Komponenten-Analyse (ICA)** für Artefaktentfernung (z. B. Augenblinzeln, Muskelartefakte), was als guter Praxisstandard gilt (vgl. Delorme & Makeig, 2004). Dabei ist zu beachten, dass bei **wenigen Kanälen (z. B. 6–8 Elektroden) die Wirksamkeit der ICA sinken kann**, da weniger Sensoren zur Trennung von Quellen zur Verfügung stehen ⁸. Mit 8 Kanälen ist ICA aber noch sinnvoll, sollte jedoch behutsam angewandt werden (ggf. Entfernung nur klarer Augenkomponenten). Die **Referenzierung** der Daten (z. B. Umreferenzieren auf Durchschnittsreferenz) stellt sicher, dass kein einzelner Elektrodenausfall die Daten verzerrt und erhöht die Vergleichbarkeit der Kanäle. Schließlich werden die kontinuierlichen Signale in **überlappende Epochen (5 s Fenster, 50 % Überlappung)** zerlegt. Eine Fensterlänge von 5 s bietet einen Kompromiss: genug Datenpunkte, um stabile Frequenzmaße (z. B. Power Spectral Density) zu berechnen, aber kurz genug, um zeitliche Veränderungen der Workload zeitaufgelöst zu verfolgen. Die 50 % Überlappung erhöht die Anzahl der Trainingsmuster und glättet zeitliche Übergänge – eine gängige Praxis, um mehr Daten aus begrenzten Versuchsdauern zu gewinnen, ohne die statistische Unabhängigkeit allzu sehr zu

verletzen (die Overlap wird bei geeigneter Validierung berücksichtigt). Insgesamt entspricht diese Preprocessing-Pipeline den **aktuellen Empfehlungen**, um artefaktbereinigte, frequenzfokussierte EEG-Daten für die Merkmalsextraktion bereitzustellen.

Modulare Pipeline-Architektur

Um Flexibilität und Wiederverwendbarkeit zu gewährleisten, wird eine **modulare Architektur** empfohlen. Die Pipeline lässt sich in klar getrennte Module unterteilen – von der Rohsignalkonditionierung bis zur Klassifikation. **Tabelle 1** gibt einen Überblick über die wichtigsten Komponenten und deren Aufgaben in der vorgeschlagenen Architektur.

Pipeline-Komponente	Aufgaben und Methoden	Best Practices
Signalvorverarbeitung	- Import der Roh-EEG-Daten (8 Kanäle, 500 Hz?) - Notch-Filter (50 Hz) & Bandpass-Filter (z. B. 1–40 Hz) - ICA für Artefaktentfernung (Augenblinzeln, EMG) - Neu-Referenzierung (z. B. auf Durchschnitt)	Entfernt Störfrequenzen und Artefakte; nutzen Sie phasenneutrale Filter (zero-phase) zur Vermeidung von Verzerrungen; bei wenigen Kanälen ICA-Komponenten manuell validieren ⁸ .
Epochierung	- Segmentierung in 5 s Epochen mit 50 % Überlappung - Label-Zuweisung pro Epoche (z. B. nach n-Back-Stufe)	Sicherstellen, dass Epochen relevante kognitive Zustände abbilden (z. B. genug Abstand nach Stimulus); Overlap berücksichtigen, um Datenlecks zwischen Train/Test zu vermeiden (bei CV).
Merkmalsextraktion	- Kanalbezogene Features: Bandpower (delta, theta, alpha, beta, gamma), Relative Power, Peak-Frequenz je Band, Band-Power-Verhältnisse (Theta/Alpha etc.) - Kanalübergreifende Features: Phasen-Kopplung (PLV) zwischen Elektroden, Kohärenz pro Frequenzband, Cross-Frequency Coupling (z. B. Theta-Phasen- zu Gamma-Amplituden-Kopplung)	Verwenden Sie bewährte Methoden: z. B. Welch-Periodogramm für PSD ⁹ , Hilbert-Transform zur Phasenberechnung für PLV. Reduzieren Sie dimensionality falls nötig (Feature Selection, PCA), um Überanpassung vorzubeugen.
Feature-Modellierung	- Zusammenführen aller Features zu Feature-Vektor pro Epoche - Optionale Normierung (z. B. Z-Score pro Kanal/Band) über Personen/Sessions - Optional: Merkmalsselektion (Entfernen redundanter Features)	Normierung hilft, personenbedingte Skalendifferenzen zu reduzieren (wichtig für Cross-Subject); Merkmalsselektion kann Rechenaufwand und Overfitting reduzieren ⁹ (z. B. Info-Gain, Korrelationskriterium).

Pipeline-Komponente	Aufgaben und Methoden	Best Practices
Klassifikation	<ul style="list-style-type: none"> - ML-Modell: Random Forest (oder Ensemble) mit hyperparam. Abstimmung (Bäume, Tiefe...)
- Training: z. B. <i>Within-Subject</i> (subjektindividuell) und <i>Cross-Subject</i> Modelle separat trainieren
- Ausgabe: Workload-Level pro Epoche (Klasse oder kontinuierlicher Score) 	Random Forests sind robust gegen Rauschen und liefern Feature-Importanzen; geeignet für kleinere EEG-Datensätze ¹⁰ . Für Cross-Subject ggf. Modell-Adaptation je Proband (Feintuning) erwägen ¹⁰ . Ergebnisse mit Kreuzvalidierung verifizieren.
Evaluation & Visualisierung	<ul style="list-style-type: none"> - Gütemaße berechnen: Accuracy, F1-Score, ggf. kappa; bei Regression ggf. RMSE
- Cross-Validation:
• Within-Subject: z. B. Train/Test-Split je Person oder CV <i>innerhalb</i> des Probanden
• Cross-Subject: Leave-One-Subject-Out (LOSO) oder Train auf n-1, Test auf neuer Person
- Ergebnis-Visualisierung: Konfusionsmatrix, ROC-Kurven (binär), Feature-Importance-Barplots, Zeitverläufe 	Strikte Trennung von Trainings- und Testpersonen für Cross-Subject-Tests, um Datenleckage zu vermeiden. Statistische Signifikanzen prüfen (z. B. gegen Zufallsbaseline). Visualisierungen helfen bei Interpretation: z. B. zeigen, welche Features dominieren (Feature Importance) oder wie Klassifikationsgenauigkeit je Proband variiert.

Tabelle 1: Modulare Architektur der EEG-Workload-Pipeline. Die einzelnen Schritte von der Rohdatenaufbereitung bis zur Ergebnisauswertung sind klar getrennt, was Wartung und Erweiterung erleichtert. Jede Komponente kann weitgehend unabhängig optimiert oder ausgetauscht werden (z. B. anderer Klassifikator), ohne die Gesamtstruktur zu gefährden.

Diese modulare Aufteilung ermöglicht es der Studierendengruppe, Arbeitsschritte aufzuteilen und jeweils **spezialisierte Methoden zu implementieren**. Beispielsweise kann das Feature-Engineering-Team neue Merkmale hinzufügen, ohne in die Preprocessing-Routine einzugreifen, solange das Datenschnittstellenformat (z. B. *epoched data*) konsistent bleibt. Ebenso kann das Modeling-Team verschiedene Klassifikatoren testen (z. B. SVM vs. Random Forest), indem sie einfach das Klassifikationsmodul austauschen. Wichtig ist, sauber definierte Schnittstellen zwischen den Modulen (z. B. Datenformate: Rohdaten -> gefilterte Daten -> Epoche-> Feature-Matrix -> Modelleingaben) festzulegen.

Feature Engineering: Merkmalsextraktion

Ziel der Merkmalsextraktion ist es, aus den 8-Kanal-EEG-Epochen informative **Features** zu gewinnen, welche Unterschiede in der mentalen Belastung widerspiegeln. Geplant ist eine Kombination aus **kanalspezifischen** Merkmalen (pro Elektrode berechnet) und **kanalübergreifenden** Merkmalen (die Wechselwirkungen zwischen Elektrodenpaaren erfassen). Im Folgenden werden die vorgesehenen Merkmalsgruppen erläutert, mit Empfehlungen für deren Berechnung im Kontext von nur 8 Kanälen (mobil) und kurzen Epochen.

1. Kanalspezifische Merkmale (pro Elektrode):

- **Absolute Bandpower:** Dies sind die *absoluten Leistungswerte* in definierten Frequenzbändern (z. B. Delta ~0.5–4 Hz, Theta ~4–8 Hz, Alpha ~8–13 Hz, Beta ~13–30 Hz, Gamma ~30–40 Hz). Die Bandpower wird klassisch über das **Power Spectral Density (PSD)**-Spektrum integriert. Eine robuste Methode ist das Welch-Periodogramm (segmentweise FFT mit Überlappung und Fensterung), um pro 5 s Epoche die mittlere Leistung je Band zu schätzen. Studien zeigen, dass **Bandpower der dominante Merkmalstyp für Workload-Klassifikation** ist ⁹. Safari et al. (2024) extrahierten z.B. band-spezifische Konnektivitätsmerkmale, aber auch dort wurden zunächst die *30 stärksten Verbindungen pro Frequenzband* ausgewählt, was den Fokus auf frequenzselektive Merkmale unterstreicht ¹¹. Für jede der 8 Elektroden erhält man pro Band einen Wert; bei z. B. 5 Bändern resultieren 40 Features. Absolute Bandpower reflektiert bekannte Workload-Effekte: **Sinkende Alpha-Power und steigende Theta-/Beta-Power** bei höherer Belastung ² ³. Diese Merkmale liefern also physiologisch interpretierbare Indikatoren des kognitiven Zustands.
- **Relative Bandpower:** Hier wird die Leistung jedes Bands auf die **Gesamtbandleistung pro Kanal** normiert (z. B. Band/Total in 1–40 Hz). Relative Power (%) beseitigt interindividuelle Unterschiede in absoluten Pegeln und langsame Drift. Insbesondere bei **personenübergreifenden Analysen** sind relative Maße oft stabiler, da sie die Variabilität der Gesamtamplitude (Kopfhautimpedanz, etc.) reduzieren. Beispielsweise könnte ein Proband generell hohe absolute μV^2 -Werte haben – relative Werte stellen sicher, dass wir wirklich die *Verteilung* der Leistung über die Frequenzen vergleichen. Methodisch: Summe der PSD über alle interessierenden Bänder pro Kanal berechnen, dann den Anteil je Band bilden. Ein Band (z. B. das Delta oder Gamma-Ende) kann man als Referenz weglassen, da die Summe aller relativen Anteile = 1 ist. Relative Bandpower ist nützlich, um z. B. Veränderungen der Alpha-Relation unabhängig vom generellen Aktivitätsniveau zu erfassen (falls z. B. in der Cafeteria generell höhere Power durch Muskulärartefakte vorliegt, bleibt der prozentuale Alpha-Anteil trotzdem aussagekräftig).
- **Peak Frequency je Band:** Dabei wird pro Band die **Frequenz mit maximaler Leistung** ermittelt. Dies kann Hinweise geben, ob sich z. B. der dominante Alpha-Peak verschiebt (etwa von 10 Hz bei niedriger Last zu 8 Hz bei hoher Last, was auf Verlangsamung hinweisen könnte) oder ob im Theta-Band ein ausgeprägter Peak auftritt (frontal oft ~6 Hz bei mentaler Anstrengung, sogenannte **frontal-midline theta**). Die Berechnung erfolgt über die PSD der Epoche: für jedes Band das Maximum suchen. Zur Glättung kann eine Interpolation oder Fit um den Peak erfolgen. Peak-Frequenzen sind empfindlich gegenüber Rauschen, daher sollte man ggf. Mindestleistungs-Schwellen definieren (um zufällige „Peaks“ im Rauschen zu vermeiden). Obwohl Peak-Frequenz in einigen Studien weniger häufig verwendet wird als Power, kann sie zusätzliche Nuancen liefern (z. B. Unterscheidung ob Alpha bei 10 Hz vs. 12 Hz dominiert). In Workload-Studien wurden z. B. Verlangsamungen im Alpha-Rhythmus unter hoher kognitiver Last beobachtet ¹², was durch solch ein Merkmal quantifizierbar wäre.
- **Band-Ratios:** Verhältniskennzahlen zwischen Bandleistungen zweier Frequenzbänder. Typische Beispiele sind **Theta/Alpha** oder **Beta/Alpha** Verhältnis. Solche Ratios compressen Informationen zweier Merkmale in eins und heben bestimmte Effekte hervor: Insbesondere Theta/Alpha wird häufig als *Workload- oder Engagement-Index* betrachtet – steigend bei höherer kognitiver Beanspruchung (Theta nimmt zu, Alpha ab) ². Ein hohes Theta/Alpha-Verhältnis deutet also auf hohe Belastung hin. Ebenso kann Beta/Alpha eingesetzt werden, da Beta tendenziell mit fokussierter mentaler Aktivität steigt ³, während Alpha sinkt. Auch Theta/Beta könnte interessant sein, z. B. in Aufmerksamkeitsstudien, doch im Workload-Kontext ist Theta/

Alpha gängiger. Die Berechnung ist trivial: Feature-Werte der betreffenden Bänder dividieren. Wichtig ist, einheitliche Definitionen zu nutzen (z.B. immer $\text{Theta}/\text{Alpha} = \text{ThetaPower}/\text{AlphaPower}$, und nicht mal invers). Band-Ratios normalisieren ähnliche Effekte wie relative Power, bieten aber ggf. stärkere Kontraste zwischen Low/High-Workload-Zuständen als einzelne Bänder isoliert.

2. Kanalübergreifende Merkmale (Elektrodenpaare):

- Phase-Locking Value (PLV):** Der *Phasenbindungswert* misst die **Konstanz der Phasenbeziehung** zwischen zwei EEG-Signalen über eine Epoche. Hierfür wird typischerweise ein engbandiges Signal pro Kanal (mittels Bandpassfilter im interessierenden Frequenzband) extrahiert und die **sofortige Phase** über die Zeit bestimmt (z. B. via Hilbert-Transformation). Der PLV ist definiert als der Betrag des **Mittelwerts der Phasendifferenz** $e^{j(\phi_1(t) - \phi_2(t))}$ über die Epochendauer. Werte nahe 1 bedeuten, dass die Phasendifferenz zwischen den zwei Elektroden über die Zeit konstant ist (starke Phasensynchronisation), Werte nahe 0 bedeuten eine variable, unsystematische Phasenbeziehung. Als Feature kann PLV für relevante Frequenzbänder und ausgewählte Elektrodenpaare berechnet werden – z.B. zwischen frontal und parietal im Theta-Band, um zu sehen ob **fronto-parietale Netzwerke** unter Last synchroner schwingen. Theoretisch könnten höhere kognitive Anforderungen *stärkere Phasenkopplung* in bestimmten Netzwerken erfordern (oder umgekehrt, Überlast könnte Kopplung stören). Die Auswahl der Paare sollte sich an Hypothesen orientieren (etwa Fz-Pz, wenn vorhanden, für frontal-posterior) oder man berechnet alle Kombinationen (8 Kanäle → 28 Paare) und reduziert später dimensionalität. Da PLV **keine Amplitudeninformation** enthält, ist es weniger anfällig für gemeinsame Artefakte (z.B. Lautstärkeschwankungen betreffen beide Kanäle ähnlich, aber Phase bleibt unabhängig). Wichtig: Um Scheinkopplung durch Volume Conduction zu mindern, kann man auch nur **Imaginary PLV** betrachten, die nur den Imaginärteil (90° Phasenversatz-Anteil) nutzt – dies erfordert aber viele Daten; bei kurzen Epochen ist Standard-PLV gebräuchlicher.
- Coherence (Kohärenz):** Kohärenz ist ein frequenzabhängiges Maß der **linearen Beziehung** zwischen zwei Signalen, definiert als normiertes Kreuzleistungsdichtespektrum. Praktisch erhält man die Kohärenz als Funktion $C(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)}$, wobei P_{xy} das Cross-PSD und P_{xx} , P_{yy} die Auto-PSDs sind. Werte nahe 1 bedeuten hohe lineare Kopplung bei Frequenz f . Für Featurezwecke kann man z.B. die **band-spezifische mittlere Kohärenz** berechnen (Kohärenz über ein Frequenzband gemittelt). Alternativ wählt man die Peak-Kohärenz in einem Band. Ähnlich wie PLV spiegelt Kohärenz die Synchronizität wider, allerdings sensitiv sowohl für Phasen- als auch Amplituden-Zusammenhänge. Kohärenz ist ein etabliertes Maß in Hirnnetzwerk-Studien; erhöhte Kohärenz in Theta oder Alpha zwischen frontalen und parietalen Ableitungen könnte auf erhöhte *funktionelle Konnektivität* bei höherer Arbeitslast hindeuten. Achtung: Laut Literatur können Kopplungsmaße variieren – z.B. fanden einige Studien, dass **unter hoher Belastung bestimmte lange Reichweiten-Kohärenzen sinken** (mögliche Aufteilung von Ressourcen), während lokale Synchronisierung steigen kann. Daher ist es sinnvoll, vorab zu prüfen, welche Verbindungen hypothesenkonform sind. Methodisch empfiehlt sich die Matlab/Python-Funktion zur Kohärenzberechnung (z.B. `matplotlib.cohere` oder `scipy.signal.coherence`), angewandt auf die 5s Epochen (evtl. mit Segmentierung innerhalb der Epoche). Ein segmentweises Mittel erhöht die Robustheit, ähnlich dem Welch-Ansatz für PSD.
- Cross-Frequency Coupling:** Dieses Feature zielt darauf ab, **Interaktionen zwischen unterschiedlichen Frequenzbändern** aufzudecken – etwa ob die Phase eines langsamen Rhythmus die Amplitude eines schnelleren Rhythmus beeinflusst. Solche *Phasen-Amplituden-*

Kopplungen (PAC) wurden mit Arbeitsgedächtnis in Verbindung gebracht, speziell die **Theta-Gamma-Kopplung** im frontalen Cortex und Hippocampus (Theta-Phase moduliert Gamma-Amplitude bei Gedächtnisbelastung) ¹³. Für die Pipeline gibt es zwei Ansätze: (a) **kanalintern** – z. B. misst man pro Elektrode, ob deren Theta-Phase systematisch mit ihrer Gamma-Amplitude korreliert; (b) **kanalübergreifend** – z. B. ob die Theta-Phase am frontalem Kanal die Alpha- oder Gamma-Amplitude an parietalem Kanal beeinflusst. Die Berechnung erfordert etwas aufwändigere Algorithmen: Ein gängiger Indikator ist der *Modulationsindex* nach Tort et al. (2010), der die Abweichung der Verteilung von Gamma-Amplituden über Theta-Phasenwinkel von der Gleichverteilung quantifiziert. Alternativ kann man analog zum PLV ein Maß definieren: z. B. für jedes Zeitfenster die Phase in Band1 und die Amplitude in Band2 bestimmen, und dann z. B. den Korrelationskoeffizienten zwischen Band1-Phase und Band2-Amplitude (oder besser: Phase -> binned amplitude mean). Toolboxes wie **BrainStorm** oder **FieldTrip** bieten PAC-Berechnungsmodule. Aufgrund der kurzen Epochen (5s) ist PAC-Schätzung allerdings nur verlässlich, wenn starke Kopplung vorhanden ist. Es bietet sich an, Theta-Gamma in frontal oder parietal zu prüfen, da n-Back-Aufgaben oft Theta-Gamma-Kopplung im **frontalen Mittellinien-Bereich** zeigen (frontal midline theta phase aligning bursts of gamma bei hoher Last). Cross-frequency Features sind fortgeschrittene Merkmale; sie können aber einzigartige Einblicke bieten: z. B. könnte ein starker Theta-Gamma-PAC ein Kennzeichen für intensive **Arbeitsspeicherbelastung** sein, während geringe Kopplung eher entspannteres Arbeiten anzeigt. Zu beachten: Bei nur 8 Kanälen und wahrscheinlich 128–256 Hz Sampling ist Gamma >40 Hz evtl. nicht sauber erfasst – man könnte Gamma auf niedriges Gamma (30–45 Hz) begrenzen oder Fokus auf Theta-Alpha Coupling legen (z. B. Theta-Phase moduliert Alpha-Amplitude, was auch in Aufmerksamkeitstudien betrachtet wird).

Die oben genannten Features sollten in einer konsistenten Struktur zusammengeführt werden – typischerweise als Feature-Vektor pro Epoche, bestehend aus allen Einzelfeatures. Dabei ist es ratsam, die Feature-Gruppen ggf. getrennt zu organisieren (z. B. erst alle Power-Features, dann alle Konnektivitäts-Features), um bei der Analyse die Möglichkeit zu haben, **Featuregruppen getrennt auszuwerten** (z. B. Leistung der Power-Features allein vs. inkl. Konnektivität vergleichen). Tabelle 2 fasst exemplarisch die wichtigsten Features zusammen:

Feature	Beschreibung & Berechnung	Nutzen / Hinweis
Bandpower (absolut)	Leistung in festem Frequenzband (Delta, Theta, Alpha, Beta, Gamma) pro Kanal; via PSD (z. B. Welch) Integrationswert ⁹ .	Hauptmerkmal in Workload-BCIs ⁹ ; Theta ↑, Alpha ↓ bei hoher Last ² . Absolutwerte können personvariabel sein (siehe relative Power).
Bandpower (relativ)	Anteil eines Bandes an der Gesamtspektralleistung pro Kanal (Bandpower / Summe aller Bänder).	Normiert individuelle Unterschiede; besser vergleichbar über Personen/Sessions. Geeignet für Cross-Subject, da Gesamtamplitudenunterschiede ausgeblendet werden.
Peak-Frequenz	Frequenz mit maximaler Power innerhalb eines Bandes pro Kanal.	Zeigt Verschiebungen dominanter Rhythmen (z. B. langsamere Alpha bei Müdigkeit/Belastung). Evtl. instabil bei kurzen Epochen – ggf. glätten.

Feature	Beschreibung & Berechnung	Nutzen / Hinweis
Band-Ratio (Verhältnis)	Verhältnis zweier Bandpower-Werte desselben Kanals, z. B. Theta/Alpha, Beta/Alpha.	Kombinationsmerkmal: verstärkt relative Änderungen. Theta/Alpha-Index steigt typ. mit Workload ² . Bietet robuste Unterscheidung zwischen Zuständen in vielen Studien.
Phasen-Locking (PLV)	Maß der Phasensynchronität zwischen zwei Kanälen in einem Band. Berechnet als Betrag des Durchschnitts der Phasendifferenzen über Zeit (0 = keine Bindung, 1 = perfekte Bindung).	Erfasst funktionelle Kopplung unabhängig von Amplitude. Kann auf relevante Paarungen (z. B. frontal-parietal Theta) beschränkt werden, um Netzwerkeffekte der kogn. Belastung zu messen.
Kohärenz (bandweise)	Frequenzspezifisches Kohärenzmaß zwischen zwei Kanälen; hier gemittelt über ein Frequenzband.	Erkennt lineare Beziehungen inkl. gemeinsamer Oszillationen. Hilfreich, um z. B. gemeinsame Alpha-Reduktion über Cortex zu quantifizieren. Mögliche Redundanz mit PLV – prüfe Korrelation beider Maße.
Cross-Frequency Coupling	Gekoppelte Maßzahl zwischen zwei Frequenzbereichen (gleicher oder unterschiedlicher Kanäle). Bsp: Theta-Phasen- zu Gamma-Amplituden-Kopplung pro Kanal.	Deckt hierarchische Interaktionen auf (z. B. Theta moduliert Gamma bei hoher kogn. Last). Könnte distinktive Indikatoren liefern, erfordert aber sorgfältige Berechnung (kurze Epochen als Limit).

Tabelle 2: Feature-Übersicht – Wichtige Merkmale zur Workload-Klassifikation mit EEG. Für jeden Merkmalstyp ist angegeben, wie er berechnet wird und weshalb er relevant ist. Literaturbefunde (rechts) stützen die Auswahl. Diese Features sollten modular generiert werden, sodass weitere Merkmale (z. B. andere Konnektivitätsmaße) künftig leicht ergänzt oder bei Unwirksamkeit weggelassen werden können.

Empfehlungen zur Feature-Extraktion: Implementieren Sie jede Featuregruppe in separaten Funktionen/Modulen (z. B. `compute_bandpower(epoch)`, `compute_plv(epoch)` etc.). So bleibt der Code übersichtlich und erweiterbar. Achten Sie auf **Parameterwahl**: z. B. Fensterlängen und Überlappung bei Welch-PSD (für 5 s Epochen evtl. 2 Segmente à 2.5 s mit 50 % Overlap innerhalb der Epoche), robuste Filter für PLV (schmalbandig, z. B. ± 1 Hz um Bandmitte, um saubere Phase zu erhalten), und ausreichende Auflösung für PAC-Berechnung (Theta-Phase typ. 4–8 Hz, Gamma-Amplitude 30–45 Hz, sicherstellen Nyquist > 90 Hz bei Samplingrate). Gegeben der begrenzten Kanalanzahl, könnte es sinnvoll sein, *a priori* nur bestimmte elektrodenspezifische Features aufzunehmen: Zum Beispiel sind bei einem 8-Kanal-Montage (vermutlich eine Auswahl aus 10-20 System) bestimmte Ableitungen entscheidend – etwa Fz, Cz, Pz, etc., je nachdem was verfügbar ist. Wenn die Montage z. B. frontal, parietal, okzipital abdeckt, liegen Workload-Effekte hauptsächlich **frontal (Theta)** und **parietal/occipital (Alpha)**. Man könnte demnach Feature-Selektion von vornherein so gestalten, dass unwichtige Kanäle in bestimmten Bändern nicht alle Kombinationen liefern (um die Merkmalsanzahl zu begrenzen). Alternativ kann nach initialem Training mit allen Features eine **automatische Merkmalsselektion** erfolgen (z. B. Information Gain oder Random-Forest-Feature-Importance, siehe nächster Abschnitt), um zu identifizieren, welche Features wirklich beitragen. Die Literatur empfiehlt,

Overfitting durch Reduktion der Merkmalsdimension zu vermeiden, gerade bei vielen konnektiven Features relativ zu wenigen Trainingsbeispielen ⁹ .

Modellierung und Klassifikation

Für die Klassifikation der extrahierten Merkmale strebt die Gruppe einen **Random-Forest-Ansatz** an. Random Forests (RF) bestehen aus einer Ensemblesammlung von Entscheidungsbäumen und gelten als **robust gegenüber Rauschen, fehlenden Werten und Overfitting** – insbesondere bei hoher Feature-Anzahl und vergleichsweise kleinem Datensatz. In EEG-BCI-Kontexten mit begrenzten Trainingsdaten haben sich RF-Modelle tatsächlich bewährt: Lotte et al. (2018) identifizieren Random Forests explizit als *besonders nützlich für Settings mit kleinem Trainingssatz* ¹⁴ . Ein RF kann nichtlineare Kombinationen von Features lernen, was hilfreich ist, da physiologische Effekte oft nicht strikt linear additiv sind. Zudem liefern RFs **Feature-Importance-Werte**, die im Nachhinein interpretiert werden können, um z. B. zu prüfen, ob bestimmte Frequenzbänder oder Konnektivitätsmaße dominieren.

Modellarchitektur: Im Rahmen der modularen Pipeline sollte das Klassifikationsmodul folgendermaßen gestaltet werden: Es nimmt einen Feature-Vektor (oder eine Feature-Matrix bei Batch-Verarbeitung) entgegen und gibt eine *Workload-Klasse* aus. Die Workload-Klassendefinition hängt vom Paradigma ab – beim n-Back könnten z. B. **drei Klassen (1-back, 2-back, 3-back)** vorliegen oder es wird vereinfachend in **niedrige vs. hohe Last** binär kategorisiert (z. B. 1-back als "low", 3-back als "high", 2-back ggf. ignoriert oder als mittlere Klasse). Die Modellierung sollte entsprechend angepasst werden (multiclass RF für >2 Klassen oder binary RF für dichotome Klassifikation).

Beim Training ist es zentral, **Within-Subject und Cross-Subject Modelle getrennt** zu betrachten:

- **Within-Subject:** Hier wird für jeden Probanden ein individuelles Modell trainiert (z. B. mit Daten der ersten Session im Labor) und auf dessen zweite Session (Cafeteria) getestet – oder per Kreuzvalidierung innerhalb der Daten dieser Person bewertet. Within-Subject hat den Vorteil, dass das Modell *personenpezifische Merkmalsausprägungen* lernen kann; typischerweise erzielt man hier höhere Genauigkeiten, da keine interindividuelle Variabilität das Modell verwirrt. In Brouwer et al.'s Studie konnte z. B. bereits nach 2.5s Daten pro Person signifikant über Zufall Workload detektiert werden ¹⁵ . Für Within-Subject empfiehlt sich eine **Cross-Validation** Strategie, z. B. *k-fold CV* auf den Epochen dieser Person (oder wenn zeitliche Abhängigkeiten relevant sind, ein blockweises CV). Auch möglich: Training auf Session1, Test auf Session2 des gleichen Probanden – um die *Session-Transfer* Stabilität zu prüfen.
- **Cross-Subject:** Hier wird ein *generisches Modell* trainiert, das für alle Probanden gelten soll. Eine rigorose Evaluierung erfolgt via **Leave-One-Subject-Out (LOSO)**: Man trainiert den Random Forest auf den Daten von 12 von 13 Probanden und testet auf dem *komplett* unbekannten 13. Probanden. Dies rotiert man, sodass jeder mal Testperson ist, und mittelt die Ergebnisse. Alternativ kann man eine fixe Aufteilung nehmen (z. B. 10 Train, 3 Test) oder ein anderes n-faches CV-Verfahren, solange sichergestellt ist, dass *keine Daten derselben Person im Training und Test vermischt* werden (sonst würden personen-spezifische Muster durchsickern und die Leistung künstlich erhöhen). Cross-Subject-Modelle sind deutlich anspruchsvoller – wie erwähnt, sanken in einem Beispiel die Genauigkeiten von ~98 % (personalisierter) auf ~74 % (generisches Modell) ⁷ . Unsere Pipeline sollte daher Mechanismen vorsehen, um Cross-Subject-Performance zu **stabilisieren**:
 - **Feature-Normalisierung über Probanden hinweg:** Ein oft notwendiger Schritt ist z. B. **Z-Score-Normalisierung pro Feature und Person** (d.h. für jeden Probanden die Feature-

Werte in Standardabweichungen relativ zum Mittelwert dieses Probanden ausdrücken). Damit werden Unterschiede im Niveau (z. B. jemand hat generell höhere Alpha-Power) entfernt, und das Modell konzentriert sich auf *relative* Unterschiede zwischen Klassen innerhalb jeder Person. Diese Normalisierung muss strikt innerhalb des Trainingsdatensatzes erfolgen – d.h. für den unbekannten Test-Probanden verwendet man die Parameter (Mittel, Std) aus dem Trainingskollektiv oder, besser, man zieht ein paar Stichproben dieses neuen Probanden zur Kalibrierung heran (semi-supervised Adaptation). Lotte et al. (2018) betonen, dass **adaptive Klassifikatoren mit (auch unsupervisierter) Anpassung an neue Nutzer deutlich überlegen** sind gegenüber statischen Modellen ¹⁰. Dies könnte bedeuten: wenn möglich, vom neuen Nutzer ein kurzes Kalibrierungsintervall aufnehmen und verwenden, um das RF-Modell nachzutrainieren oder zumindest die Feature-Skalierung zu justieren (*unsupervised*: z. B. per Distribution Matching).

- **Transfer Learning Ansätze:** In der Forschung werden auch spezifische Transferlernverfahren getestet, um die Lücke zwischen Subjekten zu überbrücken ¹⁰. Z. B. kann man das RF-Modell so erweitern, dass es *Subjekt-IDs* als Input bekommt oder Domain-Adaptation durchführt (Stichwort Riemannian alignment, siehe Barachant et al.). Für den gegebenen Rahmen kann eine pragmatische Empfehlung sein: **Trainiere ein ensembles Modell aus mehreren Random Forests**, jeweils einen pro Train-Subjekt (oder pro ähnlicher Subjektgruppe), und mittlere deren Vorhersagen – so fließen unterschiedliche individuelle Muster ein. Alternativ kann man einen einzelnen RF auf allen Trainingspersonen trainieren, aber dabei die Daten z. B. pro Person gleich gewichten (falls ungleiche Anzahl Epochen) und ggf. die Bäume tiefer wachsen lassen, um personenspezifische Zweige zu ermöglichen.
- **Regularisierung & Feature Selektion:** Ein generisches Modell profitiert davon, unwichtige oder zu stark individuenabhängige Features zu entfernen. Ein *Feature Selection*-Schritt (z. B. Info Gain Ranking ¹⁶) auf dem Trainingspool kann diejenigen Merkmale hervorheben, die consistent über Personen zwischen Klassen unterscheiden (und Merkmale, die nur bei bestimmten Individuen relevant sind, aussondern). Safari et al. (2024) kombinierten mehrere Feature-Selection-Methoden, um aus 150 initialen Features die 40 wichtigsten zu wählen, bevor sie das Modell trainierten, was die Genauigkeit auf ~89,5% an hob ¹¹ ¹⁷. Solche Strategien können gerade in Cross-Subject-Szenarien die Generalisierbarkeit erhöhen.

Modelltraining und -Validierung: Während des Trainings des Random Forest ist auf eine sorgfältige **Hyperparameter-Wahl** zu achten: Üblich sind z. B. 100–500 Bäume, maximale Tiefe unbegrenzt (RF ist relativ robust gegen Overfitting durch Mittelung), Mindestblattgrößen, etc. Hier kann eine Cross-Validation auf dem Trainingsteil zur Optimierung dienen. Wichtig: Wenn die Datengrundlage eher klein ist (13 Probanden, vermutlich jeweils einige dutzend bis wenige hundert Epochen), darf das Modell nicht zu komplex werden. RF bietet die Möglichkeit, **Out-of-Bag (OOB) Error** intern zu schätzen – das ist praktisch, um grob die Performance ohne explizite Held-Out-Set abschätzen zu können, aber letztlich sollte die eigentliche Bewertung auf den zuvor beschriebenen *unabhängigen Testpersonen* erfolgen.

Für das gegebene Experiment-Design (Labor vs. Cafeteria je Person) ergeben sich zudem interessante *Analyseoptionen*: Man könnte untersuchen, ob ein im Labor trainiertes Within-Subject-Modell auf die Cafeteria-Session derselben Person generalisiert (**Session-Transfer** robust ist). Ebenso kann geprüft werden, ob ein Cross-Subject-Modell, das auf allen Labor-Daten trainiert wurde, auf Cafeteria-Daten funktioniert und vice versa. Um Umwelt-Einflüsse (Störgeräusche, Ablenkung) zu quantifizieren, kann man die Leistung **Labor vs. Feld** vergleichen – erwartet wird tendenziell, dass im lauten Cafeteria-

Setting die Signalqualität schlechter ist und somit die Klassifikationsgüte etwas abnimmt. Dies sollte in der Evaluation berücksichtigt und ggf. berichtet werden.

Evaluation und Visualisierung der Ergebnisse

Eine gründliche **Evaluation** stellt sicher, dass die Pipeline robust und die Ergebnisse aussagekräftig sind. Zunächst sind klare *Metriken* zu definieren: Bei einer mehrstufigen Workload-Klassifikation (z. B. 3 Klassen für 1-, 2-, 3-back) bietet sich **Gesamtgenauigkeit** und **Kappa-Maß** (Korrektur für Zufallstreffer) an, zusätzlich zu Klassengenauigkeiten (Precision/Recall pro Klasse). Bei binärer Klassifikation (niedrig vs. hoch) kann man Accuracy, Sensitivity, Specificity, sowie AUC (Fläche unter ROC-Kurve) berichten. Wichtig ist, die Evaluation getrennt für Within-Subject-Modelle und Cross-Subject-Modelle durchzuführen, da die Ergebnisse hier nicht direkt vergleichbar sind. Typischerweise erwartet man *höhere Within-Subject Genauigkeiten*. Falls beispielsweise im Within-Subject-Fall ~85–90 % erreicht werden, könnten die Cross-Subject-Werte eher im Bereich 70 % liegen (angelehnt an Befunde wie in Wang et al. 2016 ~81 % ⁵ oder neueren Studien ~75 % ¹⁸). Sollte ein Cross-Subject-Ansatz deutlich schlechter als Zufall + ein bisschen (z. B. <60 % bei 2 Klassen) abschneiden, müsste man über eine Alternativstrategie nachdenken (z. B. subjektspezifische Kalibrierung zwingend erforderlich).

Kreuzvalidierung: Die gewählte Validierungsstrategie muss im Bericht transparent gemacht werden. Empfohlen wird:

- *Within-Subject:* z. B. 5-fache CV je Person (bei ausreichend Epochen). Oder "Train auf Hälfte der Epochen, Test auf anderer Hälfte" für jede Person, dann mitteln. Jede Person ergibt so eine individuelle Accuracy; diese kann man als Verteilung (Boxplot) darstellen. Brouwer et al. (2012) zeigten etwa, dass nach 2 min Daten ~85 % erreicht wurden, aber nach 2.5 s im Mittel erst ~65 % (bei Single-Trial, Fusion-Modell) – jedoch 32 von 35 Probanden signifikant besser als Zufall ¹⁵. Solche Kurven (Accuracy vs. Zeitfensterlänge) wären ebenfalls eine interessante Validierung, ob 5 s Fenster optimal sind.
- *Cross-Subject:* Leave-One-Out über 13 Probanden, wie oben beschrieben. Man erhält 13 unabhängige Test-Ergebnisse, aus denen man Mittelwert und Varianz berechnen kann. Alternativ, falls das zu varianzbehaftet ist, könnte man 3-fold CV mit je ~9 Train, 4 Test Personen (3 Permutationen) machen. Wichtig: Bei Cross-Subject kann zusätzlich *Cross-Session* getestet werden: z. B. trainiere auf allen Labor-Daten aller 13, teste auf allen Cafeteria-Daten aller 13 (so sind zwar Personen überlappend, aber Umgebungen nicht) – das prüft die Generalisierung auf neues Environment. Blum et al. (2019) berichteten z. B. ~91 % bei *cross-clip* (ähnliches Setting) aber nur ~83 % bei *cross-session* ¹⁹, was zeigt, dass Session-Wechsel (Labor->Feld) die Performance mindern kann. Solche Befunde sollte man mit der eigenen Pipeline validieren.

Statistische Absicherung: Bei 13 Probanden kann man mittels statistischer Tests prüfen, ob die mittlere Genauigkeit über Chance liegt. Dazu könnte man ein *t-Test gegen Zufallsniveau* (z. B. 50 % bei 2 Klassen) machen oder non-parametrisch durch Permutationstests die Signifikanz der Klassifikation prüfen (Labels mischen und neu trainieren, wie oft ergibt sich höhere Accuracy als echte?). Insbesondere falls neue Methoden (z. B. PLV-Features) eingeführt werden, sollte gezeigt werden, dass diese einen signifikanten Mehrwert liefern (evtl. Vergleich der Accuracy mit/ohne diese Features mittels *McNemar-Test* auf den Vorhersagen).

Ergebnisdarstellung: Für eine klare Kommunikation der Resultate sind mehrere Visualisierungen sinnvoll:

- **Konfusionsmatrix:** Zeigt für Mehrklassen-Fälle, welche Klassen häufig verwechselt werden. Etwa könnte sich zeigen, dass 2-back (mittel) teils als 1-back und 3-back fehlerklassifiziert wird, was verständlich wäre. Eine normalisierte Konfusionsmatrix (Prozente pro wahre Klasse) hilft, Stärken/Schwächen des Modells zu erkennen (z.B. perfekte 3-back Erkennung, aber Verwechslung zw. 1- und 2-back).
- **Balkendiagramm der Klassifikationsgüte pro Bedingung:** Da jeder Proband zweimal gemessen wird (Labor vs. Cafeteria), kann man die Accuracy pro Session vergleichen. Ein gruppiertes Balkendiagramm (für jeden Proband zwei Balken: Labor-Accuracy vs. Cafeteria-Accuracy) würde sofort zeigen, ob Leistung im Feld konsistent geringer ist. Auch ein Durchschnittsbalken mit Fehlerbalken für Labor vs. Feld könnte dies zusammenfassen.
- **Feature Importance Plot:** Vom trainierten Random Forest lässt sich die Bedeutung jedes Features (z.B. gemittelte Gini Gain oder Permutationsimportance) extrahieren. Ein Balkendiagramm der Top-10 wichtigsten Features kann aufschlussreich sein, um **welche Merkmale den größten Beitrag** zur Entscheidung leisten. Erwartbar wäre z.B., dass *Theta-Power frontal* und *Alpha-Power parietal* weit oben stehen, evtl. gefolgt von Theta/Alpha-Ratio etc. Wenn überraschenderweise ein exotisches Feature (z. B. PLV zwischen zwei bestimmten Kanälen) sehr wichtig ist, verdient das Diskussion. Dieser Plot stützt sich auf das trainierte Cross-Subject-Modell (oder einen über alle Train-folds gemittelten Wert). In Wang et al. (2016) wurde SVM genutzt, daher keine solchen Importances, aber andere Arbeiten mit Entscheidungsbäumen könnten hier Vergleichswerte liefern.
- **Zeitverlauf-Visualisierung:** Um die Sensitivität der Klassifikation über die Zeit zu zeigen, kann man beispielhaft für einen Probanden den *Predicted Workload Level vs. Time* aufzeichnen. Z. B. entlang des Experiments (welches vermutlich blockweise verschiedene n-Back-Levels vorgibt) kann man sehen, ob das Modell synchron mit dem Wechsel 1-back -> 2-back -> 3-back ansteigt. Dies demonstriert anschaulich die Echtzeit-Tauglichkeit. Falls der n-Back in zufälliger Reihenfolge präsentiert wurde, kann man auch die Ausgaben sortiert nach wahrer Bedingung aggregieren (z. B. mittlere vorhergesagte Wahrscheinlichkeit für "high load" während tatsächlicher 3-back Intervalle).
- **Topoplot (bei genügend Elektroden):** Mit nur 8 Kanälen ist ein Interpolations-Topoplot etwas grob, aber man könnte die räumliche Verteilung eines relevanten Features visualisieren. Beispielsweise ein Topoplot der Theta-Power-Differenz (high minus low workload) zeigt, ob erwartungsgemäß vorn größerer Anstieg ist. Oder ein Kohärenz-Netzwerk-Graph (Verbindungen zeichnen, die signifikant höher unter high workload sind). Diese Visualisierungen sind optional, können aber helfen, die neurophysiologischen Muster greifbar zu machen.

Literaturbasierte Empfehlungen: Abschließend sei betont, dass alle vorgeschlagenen Methoden in Einklang mit aktuellen Forschungserkenntnissen stehen. Die Wahl der **Spektralfeatures** als zentrale Merkmale stützt sich auf zahlreiche Studien, in denen Bandpower die zuverlässigsten Prädiktoren für mentale Beanspruchung waren ⁹. Die **Integration von Konnektivitätsmaßen** (PLV, Kohärenz) entspricht neueren Ansätzen, kognitive Zustände über neuronale Vernetzungsindikatoren zu fassen – wie etwa Safari et al. (2024), die mittels effektiver Konnektivität und Feature-Selektion eine nahezu 90%ige Klassifikationsgenauigkeit erzielten ²⁰. Bei der **Modellwahl** bestätigt der Überblick von Lotte et al. (2018) die Eignung von Random Forests für EEG-Daten mit begrenzter Größe ¹⁴, während komplexere Modelle (tiefe neuronale Netze) bislang kaum Vorteile zeigten ²¹ – ein Ergebnis, das die

Entscheidung für einen bewährten, interpretierbaren Klassifikator wie RF untermauert. Gleichzeitig unterstreicht Lotte et al., dass **adaptive und Transfer-Learning-Methoden** zukünftig wichtig sind ¹⁰ – ein Hinweis, den wir in unseren Cross-Subject-Strategien berücksichtigt haben (Normalisierung, ggf. Kalibrierung).

Zusammenfassend ist die entwickelte Pipeline – von solider Preprocessing-Routine über ein vielfältiges Feature-Set bis zum Random-Forest-Classifer – geeignet, sowohl *personenbezogen* wie auch *personenübergreifend* die mentale Workload im n-Back-Test zu klassifizieren. Mit den gegebenen Empfehlungen zu Methoden der Merkmalsextraktion und robusten Klassifikation (inkl. sorgfältiger Validierung und aussagekräftiger Visualisierung) folgt die Studierendengruppe den **aktuellen Best Practices** und Literaturerkenntnissen in diesem Bereich ¹ ² ¹⁰. Dies sollte die Erfolgchancen des Experiments mit 13 Teilnehmenden (Labor vs. Cafeteria) maximieren und zugleich eine nachvollziehbare, modular erweiterbare Forschungsplattform für zukünftige kognitive EEG-Studien schaffen.

¹ ¹⁵ Estimating workload using EEG spectral power and ERPs in the n-back task - PubMed

<https://pubmed.ncbi.nlm.nih.gov/22832068/>

² ³ ⁴ ¹² Frontiers | Working With Environmental Noise and Noise-Cancelation: A Workload Assessment With EEG and Subjective Measures

<https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2021.771533/full>

⁵ ⁸ ⁹ ¹⁶ ¹⁹ Frontiers | Monitoring pilots' mental workload in real flight conditions using multinomial logistic regression with a ridge estimator

<https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2025.1441801/full>

⁶ ⁷ ¹⁸ scitepress.org

<https://www.scitepress.org/PublishedPapers/2021/102513/102513.pdf>

¹⁰ ¹⁴ ²¹ (PDF) A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update

https://www.researchgate.net/publication/323465869_A_review_of_classification_algorithms_for_EEG-based_brain-computer_interfaces_a_10_year_update

¹¹ ¹⁷ ²⁰ Classification of mental workload using brain connectivity and machine learning on electroencephalogram data | Scientific Reports

https://www.nature.com/articles/s41598-024-59652-w?error=cookies_not_supported&code=87b17e34-1b49-44e3-a92b-f8acdc9bd7db

¹³ Frontal midline theta and cross-frequency coupling during short term ...

<https://www.sciencedirect.com/science/article/pii/S2666956022000484>