Proceedings of the
10th Graz Brain-Computer Interface Conference 2026

10.3217/978-3-99161-093-9-039

Capacity-normalized Shannon graph complexity as an EEG biomarker for event-level cognitive load estimation in adaptive BCIs

Beatriz Pascual-Roa[1,2], Eduardo Santamaría-Vázquez[1,2,3], Diego Marcos-Martínez[1,2], C. Rubén Ruiz-Gálvez[1,2], Víctor Martínez-Cagigal[1,2,3], Roberto Hornero[1,2,3]

[1]Biomedical Engineering Group, University of Valladolid, ETSIT, Valladolid, Spain
[2]CIBER-BBN, Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III, Madrid, Spain
[3]Valladolid Health Research Institute (IBioVALL), Valladolid, Spain

E-mail: beatriz.pascual@uva.es

ABSTRACT: Cognitive load (CL) reflects the mental resources required to perform a task. Adaptive brain-computer interfaces (BCIs) could benefit from EEG biomarkers that estimate CL with sufficient temporal granularity to support online adaptation. However, many EEG-based CL studies still define CL according to predefined task levels or objective task difficulty, which may not impose comparable cognitive demands across users. This assumption is especially problematic when individuals differ in working-memory capacity. We propose an event-level and capacity-normalized framework based on EEG functional connectivity in frontal and parietal regions. During the Corsi Block-Tapping Test, each encoded item was labeled according to its position relative to each participant's individual maximum working-memory capacity, allowing cognitive demand to be expressed as a percentage of individual capacity. Functional connectivity was quantified using orthogonalized amplitude envelope correlation (AEC) across theta, alpha, low- and high-beta, and gamma frequency bands, while network organization was summarized through Shannon graph complexity (S_C). The approach was evaluated in 62 cognitively healthy older adults (mean age = 70.4 ± 2.9 years). Across all frequency bands, S_C showed a progressive increase as task demands approached individual working-memory capacity. Capacity-normalized CL stages differed significantly after FDR correction ($p < 0.05$), with mostly large effect sizes ($r > 0.5$). These findings suggest that S_C, combined with capacity-normalized event-level labeling, may provide a compact and online-compatible EEG biomarker for CL-aware BCI adaptation in cognitively healthy older adults.

INTRODUCTION

Brain–computer interfaces (BCIs) perform best when they adapt to the user. This becomes challenging when the user operates under high cognitive load (CL). CL refers to the mental resources required to encode, maintain, and manipulate information during a task [1]. When task demands approach the user's limits, performance becomes unstable: errors increase and responses slow down [2]. In real-world scenarios, this can compromise safety and reliability. For this reason, closed-loop BCIs require CL estimates that are both temporally precise and individualized to each user, as the same task demand may impose different cognitive burdens depending on individual cognitive capacity and neural dynamics.

CL is often estimated with self-reports, such as NASA-TLX [3], or with behavioral measures such as accuracy and reaction time [1]. These tools are useful, but they are limited for online use. Self-reports are subjective and typically retrospective [4]. Behavioral measures provide indirect information, as they reflect performance outcomes rather than the underlying neural state [4]. In contrast, physiological signals offer a more direct window into brain activity. Electroencephalography (EEG) is particularly interesting in this context. EEG can track fast brain dynamics with high temporal resolution and can be integrated into portable BCI setups [5]. Moreover, EEG is the most commonly used signal in BCI systems, making it a natural candidate for continuous CL estimation within adaptive BCI frameworks.

Recent EEG-based CL research has increasingly explored single-trial decoding, machine-learning models, functional connectivity features, and online-compatible BCI pipelines [6, 7]. These advances show that EEG can support temporally resolved CL estimation beyond classical block-level analyses. However, a remaining challenge is how CL labels are defined. Many studies still assign CL according to objective task difficulty or predefined task levels, even though the same task level may represent low demand for one user and near-capacity demand for another. In our previous work, we showed that CL changes within a task level as items accumulate in working memory during the Corsi Block-Tapping Test (CBT) [8]. Frontal theta and parietal alpha mainly reflected general engagement, whereas beta activity tracked faster within-trial variations [8]. These findings suggest that item-level labeling can capture CL progression more precisely than block-level or level-level labels. However, item position alone is still not sufficient for cross-user

10.3217/978-3-99161-093-9-039

comparability, because the same item can represent different proportions of individual working-memory capacity.

Beyond local spectral power, CL also affects how brain regions coordinate [9]. Working memory and executive control rely on fronto-parietal interactions [10]. Functional connectivity can capture these interactions and has been increasingly used for CL estimation, including recent work based on effective or functional connectivity and machine-learning approaches [6, 11]. Graph-theoretical measures can further summarize network organization into compact features suitable for online-compatible pipelines [12]. However, connectivity-based CL studies typically still assign labels from external task levels rather than from each participant's individual cognitive capacity. Thus, combining connectivity features with capacity-normalized event-level labels may provide a more individualized representation of CL.

Two main challenges are therefore important for BCI-oriented CL biomarkers. First, features should capture CL at the time scale of cognitive events, rather than only at the level of task blocks. Second, CL labels should be normalized across users, so that CL levels reflect the proportion of individual cognitive capacity being engaged. A fixed task level does not impose the same CL on all users [13], which is especially relevant in adaptive systems where personalization is central.

The goal of this study is to evaluate whether an EEG biomarker based on network complexity can track capacity-normalized CL at the item level. To address this challenge, we propose a connectivity-based framework that combines event-level CL labeling with normalization based on each participant's Corsi span. Functional connectivity is estimated using orthogonalized amplitude envelope correlation (AEC) and summarized using Shannon graph complexity (S_C). The novelty of the proposed framework lies in combining S_C with a capacity-normalized event-level labeling strategy that expresses each encoded item as a percentage of individual working-memory capacity. This framework is evaluated in 62 cognitively healthy older adults.

MATERIALS AND METHODS

Participants and EEG recordings:

We analyzed 62 cognitively healthy older adults (46 female), aged 65-75 years (70.37 $\pm$ 2.85 years). Cognitive status was screened using the Mini-Mental State Examination (MMSE), and participants scoring below 23 points were excluded. Additional exclusion criteria included neurological or psychiatric disorders and current treatment with anxiolytic or antidepressant medication. The study was approved by the local ethics committee (PI 23-3325) and all participants provided written informed consent in accordance with the Declaration of Helsinki. EEG was recorded during the CBT implemented in the MEDUSA© ecosystem [14] using the ITACA CBT application [15]. EEG was acquired with a g.Nautilus PRO system (g.tec Medical Engineering GmbH, Austria) using 16 active hybrid electrodes at Fp1, Fp2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P3, Pz, P4, and POz (10-10 system). Conductive gel was used to improve contact quality. The common reference and ground electrodes were placed on the right and left mastoids, respectively. The sampling rate was 250 Hz. Participants followed a longitudinal protocol of nine sessions, with two CBT recordings per session.

Corsi Block-Tapping Test: CBT is a standard task to assess short-term visuospatial working memory [16]. In each trial, blocks illuminate one by one in a fixed order. Participants must reproduce the sequence in the same order. Figure 1 shows a snapshot of the digital CBT implemented within the MEDUSA© ecosystem. The encoding phase presented each item for 500 ms, with a 1000 ms interstimulus interval. Sequence length increased after correct trials, up to a maximum of nine items. In this dataset, the task started at two items. Each level included one series of two trials. A level was passed if at least one trial was correct. This incremental structure increases cognitive demands in a controlled way and allows CL to be examined at the item level as information accumulates within each trial [8].

EEG preprocessing and epoching: Preprocessing was implemented in MEDUSA© Kernel [14]. Signals were filtered with an IIR band-pass filter (0.5-60 Hz). Line noise was removed with a notch filter (49-51 Hz) [17]. We segmented EEG into 1 s epochs aligned to each encoded item. We focused on encoding because it is the stage where CL changes most within a trial [18]. To preserve compatibility with lightweight and online-compatible BCI pipelines, preprocessing was intentionally kept simple. Epochs contaminated by artifacts were automatically rejected. An epoch was discarded if the amplitude exceeded four standard deviations from the mean in at least two samples in any channel [19].

Capacity-normalized event-level labeling: We assigned CL labels at the item level, relative to each participant's visuospatial capacity. We first estimated the individual Corsi span S_i as the maximum sequence length recalled with a success ratio ≥ 0.5 across sessions [16]. This value provides a stable estimate of the participant's upper performance limit and supports cross-subject normalization.

Each trial contains a sequence of length L and item indices $k = 1, \ldots, L$. Each encoded item defines one epoch. We assigned each epoch to one of four CL stages based on the fraction k/S_i. The stages cover 0-25%, 25-50%, 50-75%, and 75-100% of individual capacity:

$$\text{Stage}(k;S_i) = \begin{cases} 1, & 0 < \frac{k}{S_i} \leq 0.25 \\ 2, & 0.25 < \frac{k}{S_i} \leq 0.50 \\ 3, & 0.50 < \frac{k}{S_i} \leq 0.75 \\ 4, & 0.75 < \frac{k}{S_i} \leq 1.00. \end{cases} \quad (1)$$

This scheme captures within-trial CL progression and makes CL labels comparable across participants with

10.3217/978-3-99161-093-9-039

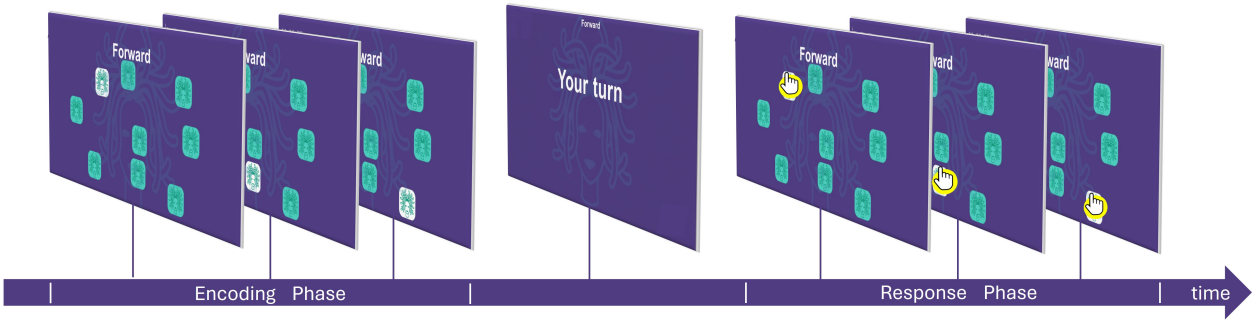

Figure 1: Snapshot of the digital Corsi Block-Tapping Test implemented in the MEDUSA© platform. During the encoding phase, blocks illuminate sequentially. After a brief interval, participants reproduce the sequence in the response phase. Sequence length increases after correct trials, progressively increasing CL.

different spans. Thus, the four stages should not be interpreted as arbitrary task phases, but as capacity-normalized CL stages ranging from low demand to near-capacity demand.

Functional connectivity and S_C: For each epoch, we estimated functional connectivity between all electrode pairs. We used AEC computed from the analytic signal after the Hilbert transform [20]. To reduce spurious zero-lag coupling associated with signal leakage and volume conduction, we used orthogonalized AEC [21]. This produced a weighted 16×16 connectivity matrix A, where A_{ij} reflects coupling strength between electrodes i and j. We summarized network organization using S_C [22]. This metric combines the dispersion of edge weights and the entropy of their distribution:

$$S_C = H\sqrt{\frac{1}{T-1}}\,\sigma_w, \tag{2}$$

where H is the Shannon entropy of the edge-weight distribution, T is the number of edges, and σ_w is the standard deviation of edge weights. We computed S_C for five frequency bands: theta (4-8 Hz), alpha (8-13 Hz), low beta (13-20 Hz), high beta (20-30 Hz), and gamma (30-60 Hz) [23]. We did not analyze delta because 1 s epochs do not capture its slow cycles reliably [24].

To obtain region-level features, we averaged node-level values within two regions of interest. The frontal region included F7, F3, Fz, F4, and F8, whereas the parietal region included P3, Pz, and P4. These regions were selected to capture broad fronto-parietal network trends associated with working memory and executive control, which are consistently involved in CL modulation [10]. The use of relatively broad ROIs was intended to prioritize robust network-level effects over hemispheric specialization in this event-level framework.

Statistical analysis: We assessed differences in S_C across the four CL stages using the Wilcoxon signed-rank test. Normality was checked with Shapiro-Wilk tests and was typically rejected. We controlled for multiple comparisons using the Benjamin-Hochberg false discovery rate (FDR) procedure with $\alpha = 0.05$ [25]. Effect sizes were computed with the rank-based correlation coefficient r, suitable for non-parametric comparisons [26]. We interpret $r > 0.5$ as a large effect and $r > 0.3$ as a medium effect [26].

RESULTS

We examined how S_C varied across the four capacity-normalized CL stages in the frontal and parietal regions. The analysis was performed separately for each frequency band. Fig. 2 summarizes these effects. The left panels show boxplots of S_C across stages for each band and region. The right panels present FDR-corrected p-value matrices and corresponding effect sizes for pairwise comparisons. Across bands, S_C increased as the encoded items approached each participant's span, with the largest increase typically occurring between the 50-75% and 75-100% stages. In contrast, the first three stages showed more gradual changes. This trend was visible in both regions and was consistent across subjects. Spatially, frontal regions generally exhibited higher S_C values than parietal regions in theta and alpha bands during low-to-intermediate CL stages (0-25%, 25-50%, and 50-75%). However, in the highest CL stage (75-100%), parietal complexity became comparable to or greater than frontal complexity in these lower-frequency bands. In contrast, low beta, high beta, and gamma bands consistently showed higher frontal complexity across all CL stages. Pairwise comparisons revealed statistically significant differences between CL stages in all frequency bands and regions after FDR correction ($p < 0.05$). In most comparisons, effect sizes were large ($r > 0.5$), indicating strong separability between CL stages.

DISCUSSION

In this study, we combined item-level CL labeling with a capacity-normalized labeling framework based on each participant's Corsi span. For every encoded item, we estimated EEG functional connectivity using orthogonalized AEC and summarized the resulting networks with S_C across canonical frequency bands.

10.3217/978-3-99161-093-9-039

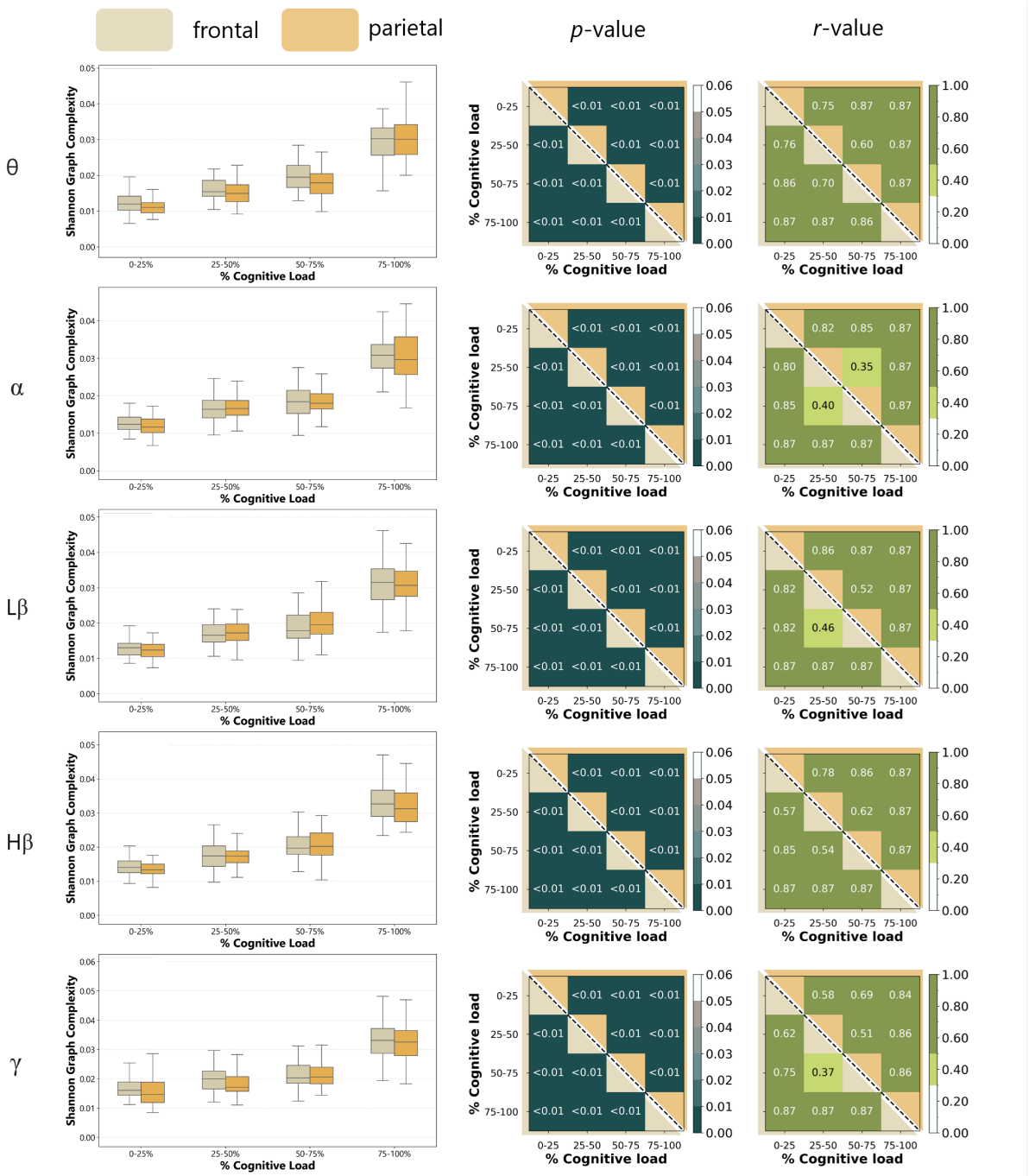

Figure 2: S_C across capacity-normalized CL stages. Each row corresponds to a frequency band: theta (4-8 Hz), alpha (8-13 Hz), low beta (13-20 Hz), high beta (20-30 Hz), and gamma (30–60 Hz). Left: compact boxplots of S_C for each stage, shown separately for frontal and parietal regions. The central line denotes the median and the box spans the interquartile range. Right: matrices of FDR-corrected p-values (left matrix) and effect sizes (r, right matrix) for pairwise stage comparisons. The lower-left triangle corresponds to frontal comparisons and the upper-right triangle to parietal comparisons. Non-white cells indicate significant differences ($p < 0.05$). Darker shades in the effect-size matrices denote larger effects.

Relation to previous EEG-based CL studies: Recent EEG-based CL research has demonstrated the feasibility of single-trial decoding, online-compatible pipelines, and connectivity-based approaches for CL estimation [6, 7]. However, many studies still define CL according to external task difficulty or predefined CL levels, which may not represent equivalent cognitive demands across users. Connectivity-based approaches using phase- and amplitude-based metrics have also been proposed [11]. In contrast, our framework assigns CL labels at the item level and anchors them to each participant's Corsi span. The resulting stages therefore reflect the proportion

10.3217/978-3-99161-093-9-039

of individual working-memory capacity being engaged, which may improve comparability across users and better match the requirements of adaptive BCIs.

Interpretation of network complexity changes: S_C summarizes the heterogeneity of connectivity weights across the network [22]. Lower values reflect a more homogeneous organization, whereas higher values indicate a broader distribution of coupling strengths. In our data, S_C progressively increased as encoded items approached each participant's working-memory span. This pattern suggests that near-capacity demands may be associated with increasingly heterogeneous functional connectivity configurations. Because S_C is a compact network-level descriptor, these findings should be interpreted as reflecting global changes in connectivity organization rather than specific edge-level mechanisms. Nevertheless, the consistent increase observed across CL stages indicates that capacity-normalized cognitive demands are accompanied by systematic changes in functional network structure. Interestingly, these effects were observed across several frequency bands, which may suggest that CL-related network reorganization is not restricted to a single oscillatory component. From a practical perspective, this broadband sensitivity could be advantageous for future online CL estimation, as it may allow flexible feature selection depending on signal quality and recording conditions. Regional differences were also observed across frequency bands. Theta and alpha bands showed relatively stronger frontal complexity during low-to-intermediate CL stages, whereas parietal complexity became more prominent near individual capacity limits. In contrast, beta- and gamma-band complexity remained predominantly frontal across stages. Although these spatial patterns should not be interpreted as direct evidence of specific neural mechanisms, they may reflect different contributions of fronto-parietal systems as cognitive demands approach individual working-memory capacity. This interpretation is consistent with the established involvement of fronto-parietal networks in working memory and executive control [10].

Implications for online-compatible CL estimation in BCIs:

Closed-loop BCIs require CL estimates that are temporally precise, stable, and individualized to each user. Our approach addresses three practical constraints. First, it produces labels at the level of individual encoded items, which is closer to the temporal scale at which adaptive BCIs may update feedback or task difficulty. Second, it includes capacity normalization, so that high CL reflects proximity to each user's own working-memory limit rather than a fixed external task level [13]. Third, S_C is a single scalar metric computed from a connectivity matrix, which keeps the pipeline lightweight. This facilitates online deployment compared with more complex dynamic approaches. In a CL-aware BCI, S_C could be used as a continuous control variable to adjust task pacing or difficulty, trigger adaptive feedback, or prevent overload during demanding interaction periods.

Limitations and next steps:

This work has several limitations. First, the distribution of epochs across stages can be imbalanced, with fewer samples near capacity. This may affect the stability of estimates in the highest CL stage. Second, the study focuses on a single task and a specific age group. Although participants were cognitively healthy older adults, mental-state properties and EEG biomarkers are known to change with age. Therefore, the present findings should be interpreted within the context of healthy aging. Future work should test the generalization to younger adults, other age ranges, and other CL paradigms. In addition, the present analysis focused on broad frontal and parietal regions to capture robust network-level trends. Future work should investigate hemispheric asymmetries and finer-grained regional effects, as lateralized neural dynamics may provide additional information about CL-related network organization. Third, we used inferential statistics but did not evaluate online classification performance. Although the proposed framework is online-compatible, the present analysis was performed offline and focused on statistical separability between capacity-normalized CL stages. Future work should evaluate online implementation, detection latency, classification robustness under realistic BCI constraints, and subject-independent models using S_C as an input feature.

CONCLUSION

In this study, we addressed the challenge of characterizing CL for BCI-oriented monitoring by leveraging EEG-derived functional brain networks under progressively increasing task demands. We introduced an event-level methodological framework that combines item-level labeling with individual capacity-based normalization and connectivity analysis. This approach allows CL progression to be described relative to each participant's working-memory capacity, rather than only according to objective task difficulty. Our results show that S_C tracks CL across canonical frequency bands in a cohort of 62 cognitively healthy older adults. As task demands approached individual capacity, frontal and parietal connectivity patterns showed progressively higher network complexity. Significant differences between capacity-normalized CL stages, together with mostly large effect sizes, support the sensitivity of this metric to graded changes in cognitive demand. These findings suggest that S_C, when combined with capacity-normalized event-level labeling, may serve as a compact EEG biomarker for CL-aware adaptive BCIs. The metric is derived from functional connectivity, summarizes network organization in a single value, and is suitable for event-level updates. Overall, this work contributes an individualized and online-compatible framework for event-level CL estimation and provides a basis for future adaptive BCI systems that respond to the user's cognitive state.

ACKNOWLEDGMENTS

10.3217/978-3-99161-093-9-039

This work was supported by '0124_EU-ROAGE_MAS_4_E project', co-financed by the EU through the Interreg VI-A Spain-Portugal Programme (POCTEP) 2021-2027, and by the VA140P24 project, funded by the Regional Government of Castilla y León (Consejería de Educación) and EU-ERDF. Support has also been provided by the Centre for Biomedical Research in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN, CB19/01/00012) through the Carlos III Health Institute, co-financed with ERDF funds.

REFERENCES

[1] Sweller J. Cognitive load during problem solving: Effects on learning. Cognitive Science. 1988;12(2):257–285.
[2] Cowan N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences. 2001;24(1):87–114.
[3] Hart SG, Staveland LE. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In: Advances in Psychology. North-Holland, 1988, 139–183.
[4] Moray N. Subjective mental workload. Human Factors. 1982;24(1):25–40.
[5] Haapalainen E, Kim S, Forlizzi JF, Dey AK. Psychophysiological measures for assessing cognitive load. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp 2010). 2010, 301–310.
[6] Safari MR et al. Classification of mental workload using brain connectivity and machine learning. Scientific Reports. 2024;14:10050.
[7] Zhou Y et al. Cognitive load recognition in simulated flight missions: Exploring temporal dynamics of multiple load states using eeg signals. Frontiers in Human Neuroscience. 2025;19:1542774.
[8] Pascual-Roa B et al. EEG biomarkers of cognitive load: Insights from incremental element encoding in short-term working memory. Biomedical Signal Processing and Control. 2026;112:108511.
[9] Ozcelik E. The neural correlates of cognitive load in learning: An fmri study on graph comprehension. Learning and Instruction. 2025;99:102175.
[10] Kim JS et al. Changes in effective connectivity according to working memory load: An fmri study of face and location working memory tasks. Psychiatry Investig.. 2012;9(3):283–292.
[11] Dong M et al. Eeg classification of resting state and arithmetic cognitive workload using functional connectivity of different frequency bands and machine learning techniques. Brain-Computer Interfaces. 2025.
[12] Liu Y et al. Fusion of spatial, temporal, and spectral EEG signatures improves multilevel cognitive load prediction. IEEE Transactions on Human-Machine Systems. 2023;53(2):357–366.
[13] Sweller J. Cognitive load theory and individual differences. Learning and Individual Differences. 2024;110:102423.
[14] Santamaría-Vázquez E et al. Medusa©: A novel python-based software ecosystem to accelerate brain-computer interface and cognitive neuroscience research. Comput. Methods Programs Biomed.. 2023;230.
[15] Marcos-Martínez D et al. Itaca: An open-source framework for neurofeedback based on brain–computer interfaces. Computational Biology and Medicine. 2023;160(March):107011.
[16] Kessels RPC, Zandvoort MJE van, Postma A, Kappelle LJ, Haan EHF de. The corsi block-tapping task: Standardization and normative data. Applied Neuropsychology. 2000;7(4):252–258.
[17] Urigüen JA, Garcia-Zapirain B. EEG artifact removal - state-of-the-art and guidelines. Journal of Neural Engineering. 2015;12(3).
[18] Pinal D, Zurrón M, Díaz F. Effects of load and maintenance duration on the time course of information encoding and retrieval in working memory: From perceptual analysis to post-categorization processes. Frontiers in Human Neuroscience. 2014;8:165.
[19] Núñez P et al. Exploring non-stationarity patterns in schizophrenia: Neural reorganization abnormalities in the alpha band. Journal of Neural Engineering. 2017;14(4):046001.
[20] Brookes MJ et al. Measuring functional connectivity using meg: Methodology and comparison with fcmri. NeuroImage. 2011;56(3):1082–1104.
[21] Colclough GL, Brookes MJ, Smith SM, Woolrich MW. How reliable are MEG resting-state connectivity metrics? NeuroImage. 2016;138:284–293.
[22] Gomez-Pilar J et al. Quantification of graph complexity based on the edge weight distribution balance: Application to brain networks. International Journal of Neural Systems. 2018;28(02):1750032.
[23] Tatum IV WO. Handbook of EEG interpretation. Analytical Biochemistry. 2014;2(1):1–5.
[24] Cohen MX. Analyzing neural time series data: Theory and practice. MIT Press: Cambridge, MA (2014).
[25] Sarkar SK. Fdr-controlling stepwise procedures and their false negatives rates. Journal of Statistical Planning and Inference. 2004;125(1-2):119–137.
[26] Pautz N, Olivier B, Steyn F. The use of nonparametric effect sizes in single study musculoskeletal physiotherapy research: A practical primer. Physical Therapy in Sport. 2018;33:117–124.