Physiological state Monitoring: a Riemannian Geometry based-model

Agathe CHOPLIN
DTIS, ONERA
13330 Salon-de-Provence, France
agathe.choplin@onera.fr

Thomas RAKOTOMAMONJY
DTIS, ONERA
13330 Salon-de-Provence, France
thomas.rakotomamonjy@onera.fr

Nicolas LANTOS
DTIS, ONERA
13330 Salon-de-Provence, France
nicolas.lantos@onera.fr

Laurent U.PERRINET
INT, Aix-Marseille Univ / CNRS
13000 Marseille, France
laurent.perrinet@univ-amu.fr

Sébastien ANGELLIAUME
DTIS, ONERA
13330 Salon-de-Provence, France
sebastien.angelliaume@onera.fr

***Abstract*— This study investigates the use of Riemannian geometry to detect and monitor physiological states such as mental workload (MWL) from an EEG dataset collected in an aeronautical context. The analysis, based on EEG data recorded from 14 participants performing a Simon's task after inducing MWL by the Multi Attribute Task Battery-II (MATB), aimed to differentiate low and high workload conditions while tracking MWL effect over time. Using covariance matrices and a Minimum Distance to Mean classifier, Temporal Generalization Method (TGM) was used to assess stable decoding performance, indicating a consistent neural signature of MWL throughout the trials. Results demonstrate spatial effects of mental workload irrespective of the investigated time domain: spatial information is distributed evenly across all explored timescale.**

Keywords*— *Mental state monitoring, classification, Riemannian geometry, electroencephalogram, spatial covariance matrices, temporal generalization, mental workload

I. INTRODUCTION

"Mental State Monitoring" (MSM) refers to the in-line or retrospective estimation of an individual's mental state using physiological and neurophysiological measurements, such as ocular activity or neuroimaging signals. The ability to reliably assess cognitive states is essential in numerous application domains, including driving assistance, aviation, and human–machine interaction [1]. In this context, passive Brain–Computer Interfaces (pBCIs) have emerged as promising tools: they acquire cerebral activity, process and classify it, and extract indicators of underlying cognitive processes [2].

Among the available neurophysiological recording modalities, electroencephalography (EEG) is particularly favored due to its low cost, non-invasive nature, and high temporal resolution, making it suitable for tracking mental states dynamically. EEG signals reflect the electrical activity generated by synchronized neuronal populations and are characterized by variations in amplitude, polarity, latency, and spatial distribution [3]. In machine learning applications, these signal properties are exploited to learn a function mapping EEG data samples to class labels within a ground-truth training dataset. Once this mapping is established, new samples can be classified, and model performance can be evaluated.

State-of-the-art machine learning approaches for EEG classification rely on the analysis of Spatial Covariance Matrices (SCMs), which condense multichannel time series by quantifying the correlations between signals collected by electrode pairs [4]. By construction, these matrices are Symmetric Positive-Definite (SPD^+) and therefore lie on the $Sym^+(n, \mathbb{R})$ manifold. This geometric structure has motivated the introduction of Riemannian geometry as a theoretically grounded framework for their processing [5]. Within this framework, two fundamental concepts emerge: distance (Affine-Invariant Riemannian distance) and mean (geometric mean) are redefined to fully exploit the structure of any SPD^+ matrix [6, 7].

Leveraging this intrinsic geometry enables a more faithful representation of manifold curvature and inter-matrix distance, ultimately leading to improved classification performance [6, 8]. Riemannian-based models have been successfully applied to various neurocognitive decoding tasks, including the characterization of motor imagery [5, 8] and the monitoring of mental states [9].

Mental workload (MWL) refers to the cognitive demand imposed by a task and the mental effort required to perform it. However, these resources are inherently limited: when task demands exceed the available capacity, cognitive overload occurs, often leading to performance degradation [10]. An increase of cognitive load can result in a modulation of brain activity measurable using EEG [11].

This work aims to investigate time-resolved decoding of a physiological state, here MWL, and the model ability to generalize over time using Riemannian classification and Temporal Generalization methodology [12]. Segmenting signals into sub-windows were used to provide a framework for investigating the variability and temporal dynamics of the spatial signature of MWL in EEG signal, enabling a precise assessment of the behavior of the model and performances with the distribution of data evolving over time. The model functions as a proxy measure aiming to enable the monitoring of physiological state in real-world applications.

II. MATERIAL

The present study relies on a previously acquired dataset investigating MWL within an aeronautical context, collected by ONERA (the French Aerospace Lab) [11]. EEG activity was recorded from 21 participants exposed to different experimentally manipulated workload conditions, using 64-channel active-electrodes positioned according to the international 10|20 system and sampled at 500 Hz.

979-8-3315-7927-2/26/$31.00 ©2026 IEEE

Mental workload was measured using the Multi-Attribute Task Battery II (MATB-II) [13], configured with two difficulty levels—low and high—completed in two separate sessions. Workload levels were assessed by adjusting both the number of sub-task and their execution speed. Following each MATB-II session, participants completed three evaluation tasks. Those tasks were selected to be representative of the three main executive functions, involved in piloting activities: inhibitory control (Simon task [14]), cognitive flexibility (Switching task [15]), and working memory (N-back [16]). This design enabled the assessment of cognitive function influenced by the prior workload manipulation. This paper is focusing only on the results of the Simon task.

III. METHOD

A. Analysis

EEG recordings from 14 participants were retained for the subsequent analyses: 7 of the original 21 subjects were excluded due to artifacts rejection during preprocessing stage. For each remaining participant, EEG data were preprocessed by interpolating bad channels, applying a 50 Hz notch filter, a 0.1 Hz high-pass filter, and re-referencing signals to the common average. Artifact removal was performed using Signal-Space Projection (SSP) [17]. Ocular artifacts were identified through dedicated EOG channels for blink detection and AF7|AF8 electrodes for saccade detection.

Stimulus-locked epochs were extracted for all the EEG data, with a baseline correction applied over the interval [-0.1s, 0s], relative to stimulus onset at 0s and a total epoch interval of [-0.3s, +1.0s]. For each participant and workload condition, around 273 (± 20) stimulus-locked epochs were created for each recorded signal depending on the previous artifact rejection phase. Based on previous findings [11] evidencing a strong workload-related effect over central electrodes (P_z, CP_z, C_z, FC_z, F_z), the analysis focused on these sites, supplemented by five additional electrodes on each lateral scalp region (right: P_2, CP_2, C_2, FC_2, F_2; left: P_1, CP_1, C_1, FC_1, F_1). According to the results reported in [18], using a smaller set of electrodes to compute Spatial Covariance Matrices (SCMs) reduces computational cost, while preserving good classification performance.

The study aim was to classify mental states associated with low versus high workload across EEG time course, thereby enabling a continuous monitoring of workload dynamics. The following section will describe the proposed methodology. The algorithm was implemented using Python and libraries such as *MNE* (v.1.11.1) for EEG data processing, *pyRiemann* (v.0.9) for Riemannian geometry computations, and *scikit-learn* (v.1.7.2) for machine learning procedure.

B. Methodology

Previous work [19] has demonstrated a high level of classification performance on the ERP epochs of this dataset using Riemannian manifold–based methods. Building on this, the present study aims to determine whether specific segments of the signal contain discriminative information, and to identify those intervals that most strongly contribute to classification performance.

Each stimulus-locked epoch ([-0.3s, +1.0s]) was first segmented into 0.1s sub-windows with 50% temporal overlap between consecutive windows, resulting in a total set of 25 sub-windows per epoch. As SCMs do not account for temporal structure, segmenting the data into shorter epochs enables the extraction of temporal dynamics within the EEG signals. For every of these sub-windows, SCMs were estimated using the Oracle Approximated Shrinkage (OAS) method [20], yielding in a set of 25 SCMs per epoch. These matrices constituted the input features for classification with the following Riemannian classification method.

Within-subject classification was performed using the Minimum Distance to Mean (MDM) classifier [7] based on Riemannian's metrics. The Affine-Invariant Riemannian distance (AIR) (1) quantifies the shorter length (e.g., the geodesic) between two SPD+ matrices, Σ_1 and Σ_2, on the manifold. With $\|.\|F$ the Frobenius norm of a matrix.

$$\delta_R(\Sigma_1, \Sigma_2) = ||log(\Sigma_1^{-1}\Sigma_2)||_F = \sum_{c=1}^{C} log^2\lambda_n, \qquad (1)$$

The geometric mean (2) corresponds to the barycenter of a set of matrice on the manifold, i.e., the matrix minimizing the sum of squared Riemannian distances to all matrices of the dataset:

$$\bar{C}(\Sigma_1, \dots, \Sigma_I) = argmin_{\Sigma \in \mathcal{P}(n)} \sum_{i=1}^{I} \delta_R^2(\Sigma, \Sigma_i) \qquad (2)$$

The MDM classifier operates on two steps; First, during the learning phase the model estimates barycenter of each class, i.e., the mean covariance matrix of the associated trials for both low- and high-workload conditions. Then during the testing phase, for each new trial, the distance to each class barycenter is computed, the trial is assigned to the class whose Riemannian distance from the barycenter is the smallest.

Model performance was assessed through a shuffle-split cross-validation scheme consisting of 20 iterations, each employing 80% of the data for training. For each split, the model was trained on all the epochs of the training set for a given time window and tested on all epochs of the test set across all time windows, thereby generating a Temporal Generalization Matrix (TGM) [12]. We use the classification accuracy as a proxy to determine whether the information learned at a given time point *t* allows for decoding at another time *t'*.

IV. RESULTS

A. Classification scores results

The classification of mental workload performed with the MDM classifier and a cross-validation scheme (20 splits with 80% used for training), based on the full stimulus-locked epoch, achieved a mean accuracy level on subjects (N=14) of 0.95 with a standard deviation of 0.03.

Additionally, classification performance was assessed across time sub-windows (0.1s sub-windows with 50% temporal overlap) and cross-validation folds (20 splits with 80% used for training). We obtain a mean inter-subject (N=14) accuracy over the sub-windows of 0.83 with a standard deviation of 0.01.

B. Temporal decoding results

For the temporal decoding results, we first analysed the classification scores obtained using the same temporal sub-window across all epochs for training and testing (Fig. 1). The mean accuracy on participants (N=14) and cross-validation splits (20 folds, 80% for training) over time reach a classification score of 0.83 (± 0.01). In addition to the average value, we also plot the standard deviation (SD) of the

979-8-3315-7927-2/26/$31.00 ©2026 IEEE

classification scores obtained through the folds. The temporal analysis reveals a significant decoding regardless the selected time interval and in particular for pre-stimulus periods. We also notice a small increase of accuracy around +0.20s post-stimulus (maximal accuracy= 0.85).

C. Temporal Generalization results

The Fig. 2 (A) gives the mean temporal generalization matrix on all the participants (N=14) and folds (20 splits with 80% used for training), obtained when mixing the selected sub-intervals for training and testing. This matrix is a sub-window by sub-window matrix (25 x 25), the x-axis are the tested windows and the y-axis is representing the time of trained windows. The vertical and horizontal black lines represent the stimulus onset (0s).

The results of the main diagonal are reported in the previous sub-section (Fig. 1). The off-diagonal results indicate a sustained and generalized effect on all sub-windows times, with high accuracy scores (maximal accuracy= 0.85; minimal accuracy= 0.80).

D. Statistics on Temporal Generalization results

This analysis aim is to determine whether the differences observed between the pixels of the generalization matrix are statistically significant, in order to assess whether specific train–test intervals yield optimal performance or whether classification accuracy remains relatively uniform across the matrix. We extracted the maximum and minimum value (min mean matrix = 0.80 and argmin= [+0.80, +0.35]s, max mean matrix = 0.85 and argmax= [+0.20, +0.20]s) of the mean matrix across-subjects.

The differences distribution was tested for normality (Shapiro–Wilk: W = 0.834, p = 0.0138). In the absence of normality, we applied a Wilcoxon signed-rank test: V = [11.0], p = 0.015, r = [0.64] (Fig. 3). The 95% confidence intervals for the mean difference were [0.017; 0.110].

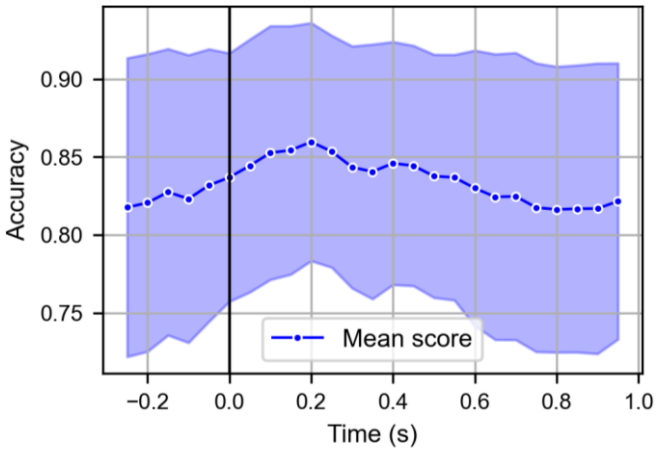

Fig. 1. Temporal decoding curve. Mean curve of accuracy classification on all the participants (N=14) and folds (20 split and 80% for train) over time, blue line (mean) and band ± SD.

V. Discussion

Application of Riemannian geometry framework to EEG covariance matrices yielded high classification performance (0.95 ± 0.03) on our dataset. Previous work [21] reported an accuracy of 0.64 for binary mental workload classification using a Support Vector Machine (SVM) classifier trained on a single feature. This result is consistent with accumulating reference demonstrating the superiority of Riemannian method for the classification of physiological and cognitive states [6, 8].

Temporal decoding analyses (Fig. 1) revealed unexpectedly strong classification performance during the baseline interval, reaching levels comparable to those observed after stimulus presentation. Moreover, decoding accuracy remained relatively stable across different time windows, with a maximum score located around +0.20s. In light with previous ERPs analyses on this dataset [11], this peak could be plausibly attributed to a transient increase in of Signal-to-Noise Ratio (SNR) at that latency. Those findings

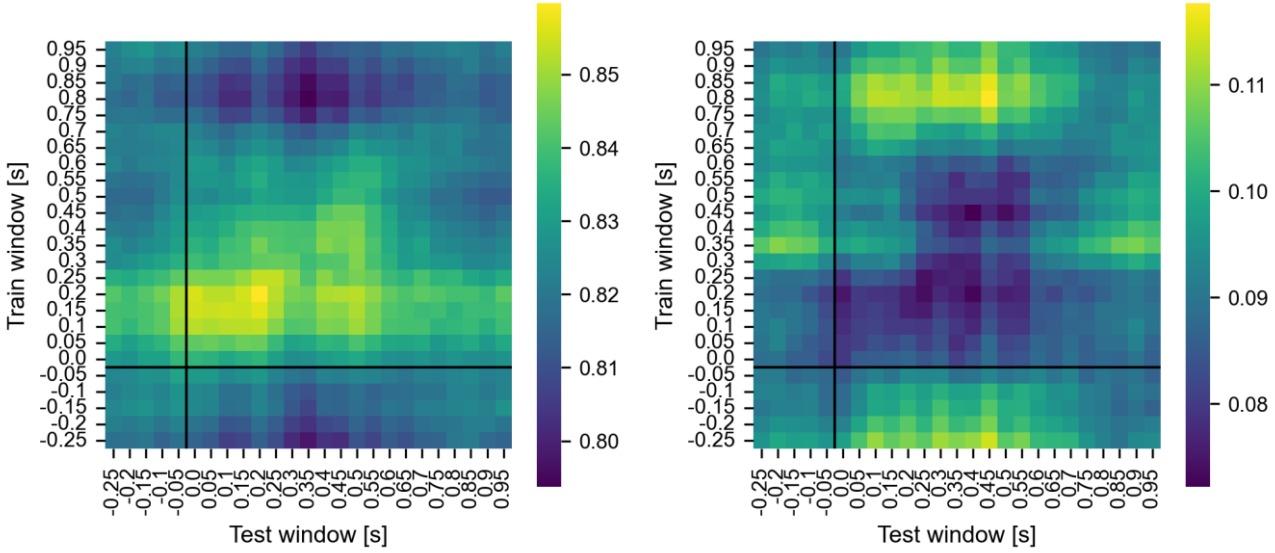

Fig. 2. Temporal generalization matrice of MWL decoding. (A) Mean inter-subject and folds TGM ; (B) Standard Deviation inter-subject and folds TGM. Each cell (i,j) represent the accuracy when the model is trained on the window i (y axis) and tested on the j window (x axis). Black lines = stimulus onset at 0s. The diagonal pixels of the matrix are the specific "temporal decoding". The Off-diagonal pixels of the matrix are the temporal generalisation.

979-8-3315-7927-2/26/$31.00 ©2026 IEEE

demonstrate that the spatial signature of the workload effect in the brain manifests itself over the epoch domain.

The temporal generalization matrix (Fig. 2A) further supports this interpretation. A coherent region of high temporal generalization emerged between roughly [0s, +0.25s], for both training and testing time points. The diagonal—together with sustained off-diagonal performance—suggests the presence of temporally stable neural representations that support generalization across distinct timescales.

Inter-participant variability analyses (Fig. 2B) further demonstrate that this temporal interval (around 0s, +0.25s) is associated with the smaller across-subject variability. Standard deviation values were minimal within the [0s, +0.25s] range, indicating that the neural dynamics underpinning the decodable information are robust but also reliably conserved across individuals.

Statistical comparison analysis between the highest and lowest values of the TGM has shown significant differences ($p < 0.0159$) with a high effect size ($r = 0.64$) (Fig. 3). These results suggest a significant difference, indicating that certain train–test combinations (maximum value = 0.85) achieve substantially higher performance than others (minimum value = 0.80), rather than reflecting a uniformly distributed classifier's accuracy across the entire matrix. Further statistics analysis such as permutation-cluster-based test [22] could be applied to the mean of TGM to ensure this statistical significance.

The methodological framework introduced here is not restricted to MWL estimation. It generalizes naturally to transient cognitive or affective states such as acute stress or surprise, and can accommodate diverse physiological modalities (e.g., electro-ocular, or electro-cardiac activity). It also lends itself to multimodal integration within hybrid passive BCIs, for which the combination of complementary metrics has been shown to enhance classification performance [23].

Finally, we followed analysis by applying this approach in a multitasking context— by training the model on two tasks data (e.g., Simon, N-Back) and evaluating it on a third task (e.g., MTAS) —yielded strong cross-task generalization, with a mean accuracy of 0.78 ± 0.1.

This study demonstrates that, even within ERP data, cognitive states can be assessed without confining the analysis to specific components such as the P300. Instead, the relevant information appears to be distributed across the entire epoch domain. Within this framework, such approach offers promising perspective for detection and tracking of cognitive states (e.g., stress, overload, surprise), that may be implemented in real-time devices to provide early alerts and decision-support clues to enhance safety and performance in operational settings such as aviation.

ACKNOWLEDGMENT

The authors would like to thank Office National d'Etude et de Recherche en Aérospatial and the Région Provence-Alpes Côte d'Azur for funding the Ph.D. thesis of Agathe Choplin. We gratefully acknowledge Deshayes C., Berberian B., and Ficarella S.C., for providing access to their dataset and granting us permission to utilize it in our study.

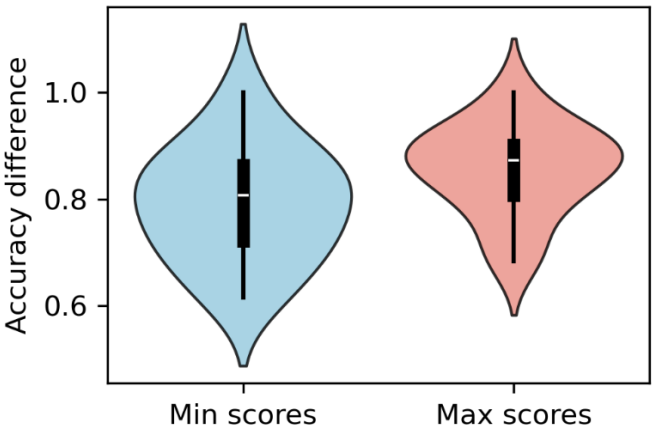

Fig. 3. Distribution of minimum and maximum accuracy scores among participants. The inner box indicates the median and quartiles. The Wilcoxon test shows that maximum accuracy is significantly higher than minimum accuracy ($W = 11$, $p = 0.016$, $r = 0.64$).

REFERENCES

[1] R. Parasuraman, J. Christensen, and S. Grafton, "Neuroergonomics: The brain in action and at work," NeuroImage, vol. 59, no. 1, pp. 1-3, 2012, doi: 10.1016/j.neuroimage.2011.08.011.

[2] L. George and A. Lécuyer, "An overview of research on 'passive' brain-computer interfaces for implicit human-computer interaction," 2010. [Online]. Available: https://hal.inria.fr/inria-00537211/document

[3] M. X. Cohen, *Analyzing Neural Time Series Data: Theory and Practice*, 2014. [Online]. Available: https://doi.org/10.7551/mitpress/9609.001.0001

[4] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," Neurocomputing, vol. 112, pp. 172–178, 2013, doi: 10.1016/j.neucom.2012.12.039.

[5] M. Congedo, A. Barachant, and A. Andreev, "A new generation of Brain-Computer interface based on Riemannian geometry," HAL, 2013. [Online]. Available: https://hal.science/hal-00879050

[6] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review," 2017.

[7] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Riemannian geometry applied to BCI classification," in *Lecture Notes in Computer Science*, 2010, pp. 629–636, doi: 10.1007/978-3-642-15995-4_78.

[8] F. Yger, M. Berar, and F. Lotte, "Riemannian Approaches in Brain-Computer Interfaces: A Review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1753–1762, 2017, doi: 10.1109/TNSRE.2016.2627016.

[9] S. C. Wriessnegger, P. Raggam, K. Kostoglou, and G. R. Müller-Putz, "Mental state detection using Riemannian geometry on electroencephalogram brain signals," Frontiers in Human Neuroscience, vol. 15, 746081, 2021, doi: 10.3389/fnhum.2021.746081.

[10] B. Hsu, M. J. Wang, C. Chen, and F. Chen, "Effective indices for monitoring mental workload while performing multiple tasks," *Percept. Mot. Skills*, vol. 121, no. 1, pp. 94–117, 2015, doi: 10.2466/22.pms.121c12x5.

[11] C. Deshayes, S. Angelliaume, B. Berberian, and S. C. Ficarella, "The quest for a task-independent (neuro) physiological signature of cognitive fatigue," *Manuscript in preparation*.

[12] J. R. King and S. Dehaene, "Characterizing the dynamics of mental representations: the temporal generalization method," *Trends Cogn. Sci.*, vol. 18, no. 4, pp. 203–210, 2014, doi: 10.1016/j.tics.2014.01.002.

[13] Y. Santiago-Espada, R. R. Myer, K. A. Latorella, and J. R. Comstock, *The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User's Guide*, 2011. [Online]. Available: https://ntrs.nasa.gov/citations/201100144

[14] J. L. Craft and J. R. Simon, "Processing symbolic information from a visual display: interference from an irrelevant directional cue," *J. Exp. Psychol.*, vol. 83, no. 3, Pt. 1, pp. 415–420, 1970, doi: 10.1037/h0028843.

[15] A. T. Jersild, “Mental set and shift,” *Archives of Psychology*, vol. 89, pp. 5–82, 1927.

[16] S. M. Jaeggi, M. Buschkuehl, W. J. Perrig, and B. Meier, “The concurrent validity of the N-back task as a working memory measure,” *Memory*, vol. 18, pp. 394–412, 2010.

[17] M. A. Uusitalo and R. J. Ilmoniemi, "Signal-space projection method for separating MEG or EEG into components," *Med. Biol. Eng. Comput.*, vol. 35, no. 2, pp. 135–140, 1997, doi: 10.1007/bf02534144.

[18] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Common Spatial Pattern revisited by Riemannian geometry," in *IEEE Int. Workshop on Multimedia Signal Processing*, 2010, pp. 472–476, doi: 10.1109/mmsp.2010.5662067.

[19] A. Choplin, T. Rakotomamonjy, L. Perrinet, N. Lantos, S. Angelliaume, “Classification of Mental Workload Frequential Effects using Riemannian Manifold,” poster presented at *Cognitive Computationnal Neurosciences*, Amsterdam, NL, August 2025

[20] Y. Chen, A. Wiesel, Y. C. Eldar, and A. Hero, "Shrinkage algorithms for MMSE covariance estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5016–5029, 2010, doi: 10.1109/tsp.2010.2053029.

[21] A.-M. Brouwer, M. A. Hogervorst, J. B. F. van Erp, T. He elaar, P. H. Zimmerman, and R. Oostenveld, "Estimating workload using EEG spectral power and ERPs in the n-back task," *J. Neural Eng.*, vol. 9, no. 4, 2012.

[22] E. Maris and R. Oostenveld, “Nonparametric statistical testing of EEG- and MEG-data,” *Journal of Neuroscience Methods*, vol. 164, no. 1, pp. 177–190, 2007, doi: 10.1016/j.jneumeth.2007.03.024.

[23] K. Ryu and R. Myung, "Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic," *Int. J. Ind. Ergon.*, vol. 35, pp. 991–1009, Nov. 2005.

979-8-3315-7927-2/26/$31.00 ©2026 IEEE