ARL-TR-10476 ● Oct 2026



DEVCOM ARMY RESEARCH LABORATORY

Effects of EEG Preprocessing Methods on Modified Diffusion Entropy Analysis and Complexity Synchronization

by Sabrina Sullivan, Scott E. Kerick, and J. Cortney Bradford

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-10476 ● OCT 2026





Effects of EEG Preprocessing Methods on Modified Diffusion Entropy Analysis and Complexity Synchronization

Sabrina Sullivan
Oak Ridge Associated Universities

Scott E. Kerick and J. Cortney Bradford
DEVCOM Army Research Laboratory

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
October 2026	Technical Report	January 6, 2025	February 26, 2026

4. TITLE AND SUBTITLE
Effects of EEG Preprocessing Methods on Modified Diffusion Entropy Analysis and Complexity Synchronization

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
W911NF-22-2-0097 (ORAU)		
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Sabrina Sullivan, Scott E. Kerick, and J. Cortney Bradford

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
DEVCOM Army Research Laboratory ATTN: FCDD-RLA-IJ Aberdeen Proving Ground, MD 21005	ARL-TR-10476

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
ORCIDs: Sabrina Sullivan, 0000-0001-8660-0311; J. Cortney Bradford, 0000-0002-5926-9229; Scott E. Kerick, 0000-0001-6177-1552

14. ABSTRACT
Electroencephalogram (EEG)-derived complexity metrics offer insight into the brain's capacity for adaptive information processing; however, the influence of artifact removal on these estimates remains poorly understood. This study investigated how EEG preprocessing and artifact removal affect estimates of neural complexity and complexity synchronization (CS), as quantified by modified diffusion entropy analysis (MDEA), a method grounded in crucial event theory. Two EEG datasets were analyzed: one dominated by persistent artifacts from treadmill walking with exoskeleton assistance, and one dominated by transient artifacts from a seated high- and low-stress shooting task. MDEA was applied to the datasets after each of the four preprocessing stages: 1) robust re-referencing with high-pass filtering, 2) artifact subspace reconstruction, 3) canonical correlation analysis, and 4) adaptive mixture independent component analysis (ICA) with IC pruning. In both persistent and transient artifact datasets, mean and standard deviation of complexity estimates decreased from minimally to maximally processed data; however, temporal fluctuations in complexity were largely preserved across all preprocessing stages. The relationship between conditions remained intact throughout the pipeline for the persistent-artifact dataset (pre- vs. post-adaptation). The CS between EEG and electrocardiogram was consistently significant during low- and high-stress conditions but it was consistently greater during low stress. These results indicate that although artifact removal systematically lowers the magnitude of complexity estimates, the temporal structure underlying neural complexity is largely robust to artifact contamination, suggesting that fewer preprocessing steps may suffice for some applications and offer practical benefits for computationally efficient, real-time neural monitoring.

15. SUBJECT TERMS
electroencephalogram, EEG, artifact, complexity, synchrony, Humans in Complex Systems

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	30

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
J. Cortney Bradford	(520) 691-6370

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures v

Acknowledgment vi

1. Introduction 1

1.1 EEG Artifacts and Removal Approaches 1

1.2 Neural Complexity, Synchrony, and System Adaptation 2

2. Objectives 4

3. Methods 4

3.1 Datasets 4

3.2 EEG Preprocessing Pipeline 5

3.2.1 Minimal Preprocessing 6

3.2.2 Artifact Subspace Reconstruction 6

3.2.3 Canonical Correlation Analysis 6

3.2.4 AMICA with IC Pruning (Maximally Processed) 6

3.3 Modified Diffusion Entropy Analysis 7

3.4 Complexity Synchronization 9

4. Results 9

4.1 Persistent Artifact Study 9

4.2 Transient Artifact Study 10

5. Discussion 11

5.1 Reduction in Neural Complexity and Variability with Artifact Removal 12

5.2 Neural and Organ System Synchrony and CS 14

6. Limitations 14

7. Conclusion 15

8. References **16**

List of Symbols, Abbreviations, and Acronyms **21**

Distribution List **22**

List of Figures

Figure 1. EEG preprocessing pipeline. The left panel shows the sequential preprocessing steps and corresponding EEGLAB/MATLAB functions used in the pipeline. The right panel shows an example of the same EEG signal segment after each successive cleaning stage, illustrating how the signal changes across preprocessing steps. 5

Figure 2. MDEA on data with persistent artifacts. a) An example of one subject for one condition where temporal fluctuations are preserved through all processing steps in Pre-adaptation. b) A Spearman rank correlation heat matrix annotating the Fisher-z corrected averaged Spearman rho between all EEG processing pipeline steps in Pre- (blue) and Post- (red) adaptation averaged across subjects. c) A bar plot comparing the mean complexities (δ) across the EEG cleaning pipeline over time. Statistical significance is denoted by an asterisk: *p < 0.05, **p < 0.01, and ***p < 0.001. 10

Figure 3. MDEA on data with spurious artifacts. a) An example of one subject whose temporal fluctuations are preserved through all processing steps in high stress, with an asterisk denoting significant decrease in mean scaling. b) A Spearman rank correlation heat matrix annotating the Fisher-z corrected averaged Spearman rho between all EEG processing pipeline steps in low (blue) and high (red) stress averaged across subjects. c) A bar plot comparing the mean complexities (δ) across the EEG cleaning pipeline over time. d) A Spearman correlation heat matrix annotating the Fisher-z corrected averaged Spearman rho between EEG and ECG scaling indices across each cleaning step in the processing pipeline. Statistical significance is denoted by an asterisk: *p < 0.05 and **p < 0.001. 11

Acknowledgment

The authors acknowledge Drs. Korosh Mahmoodi and Piotr Franaszczuk for valuable feedback and comments on the manuscript.

1. Introduction

Electroencephalogram (EEG)-derived complexity metrics have advanced substantially over the past three decades, progressing from approximate entropy to sample entropy, in addition to more recent approaches such as multiscale entropy, permutation entropy, and modified diffusion entropy, which address known limitations including self-matching bias and estimator instability.[1] Despite these methodological improvements, it is uncertain how much non-neural artifacts and their removal influence the estimation of neural complexity. Ideally, EEG complexity should index the brain's capacity for flexible information exchange and adaptive responses to cognitive and environmental demands or changes.[2–4] EEG recordings are unavoidably affected by artifacts arising from muscle activity, environmental noise, electrode disturbances, instrumentation, and systemic physiological rhythms such as cardiac and respiratory cycles.[5] These artifacts can be persistent, such as line noise, or transient, such as sporadic eye movements, and both may bias complexity estimates. Artifact removal techniques including artifact subspace reconstruction (ASR), canonical correlation analysis (CCA), and adaptive mixture independent component analysis (AMICA) can improve signal quality and attempt to isolate neural activity,[5] but they may also alter neural signal structure in ways that complicate interpretation. As a result, EEG complexity estimates must be evaluated with careful consideration of preprocessing choices and their potential impact on the underlying neural signal.

In this study, we aimed to understand how EEG signal preprocessing approaches impact the estimation of neural signal complexity using modified diffusion entropy analysis (MDEA). We used two existing EEG datasets, which contained either primarily persistent or transient artifacts, to represent a range of artifact types.

1.1 EEG Artifacts and Removal Approaches

Artifacts in electroencephalogram recordings are signals that do not originate from the brain and are broadly categorized as either physiological, stemming from the subject's body, or non-physiological, arising from external sources such as equipment.[6,7] It is crucial to identify these artifacts because they can obscure or be mistaken for genuine brain activity, potentially leading to misinterpretation. Common physiological artifacts include ocular artifacts from eye movements and blinks; high-frequency muscle (myogenic) artifacts from activities like chewing or clenching the jaw; cardiac (ECG) artifacts synchronized with the heartbeat; glossokinetic artifacts caused by tongue movement; and slow, undulating sweat or respiration artifacts resulting from perspiration and breathing.[6,8] Non-physiological artifacts, originating from outside the body, encompass electrode "pops" from loose

connections, persistent interference from electrical power lines, and various movement artifacts generated by physical disturbances such as during walking.[6,9,10]

To address the challenge of artifacts in EEG, a variety of powerful tools have been developed to either remove or significantly reduce their impact.[11] These methods often target one or more specific types of artifacts, using different strategies to distinguish and separate the unwanted noise from the underlying brain activity. Basic filtering techniques, such as band-pass filtering, are used to isolate specific frequency bands of interest while removing low-frequency drifts and high frequencies often associated with muscle artifacts.[5] More advanced methods leverage blind source separation techniques, which attempt to statistically separate the mixed signals recorded at the scalp into their original sources. Independent component analysis (ICA) is a widely used method that can effectively identify and remove various artifacts, including eye blinks, muscle activity, and cardiac signals, by isolating them into independent components (ICs) that can then be subtracted from the data.[11,12] Another advanced technique is ASR, which is an automated, component-based method capable of removing large-amplitude, transient artifacts by identifying and reconstructing contaminated data segments.[13] For the specific challenge of muscle artifacts, which often overlap in frequency with neural signals, methods like CCA[14] and a combination of ensemble empirical mode decomposition with CCA have proven to be effective.[15,16] Empirical mode decomposition itself is another adaptive data analysis method that can decompose a signal into a set of intrinsic mode functions, which can then be analyzed to identify and remove artifactual components.[15] These diverse approaches, used alone or in combination, provide researchers with a robust toolkit for improving the quality and reliability of EEG data.

A major challenge in EEG artifact removal is balancing artifact suppression with preservation of neural signals.[11] For example, ASR excels at detecting high-variance outliers but may unintentionally suppress actual neural variability,[13] whereas AMICA, effective at isolating independent sources, can overfit data.[17] EEG complexity estimation can be affected by these preprocessing choices. Over-aggressive artifact removal can artificially flatten complexity measures, whereas insufficient cleaning allows residual noise to obscure meaningful patterns.[11,18]

1.2 Neural Complexity, Synchrony, and System Adaptation

Neural complexity, measured from electroencephalogram recordings, is a valuable tool for understanding brain function in both health and disease. It quantifies the 1/f-variability and unpredictability of brain signals, which can reflect the brain's information processing capacity. Complexity is hypothesized to represent the

brain's ability to generate diverse and adaptive responses. It captures the dynamic balance between integration and segregation across neural networks.[19] This balance supports the brain's capacity for flexible adaptation to internal and external demands and perturbations, supporting cognitive resilience and system performance.[20] For example, resting-state EEG complexity has been shown to predict individual differences in cognitive performance, suggesting that richer neural dynamics may support more effective information processing.[21] In contrast, reductions in neural complexity have been observed under conditions of cognitive strain or fatigue.[22] Zhai et al.[22] reported that mental fatigue is associated with decreased EEG complexity, consistent with the theory that demanding cognitive activity can reduce the dynamical variability of neural signals. More broadly, EEG complexity measures, such as entropy, fractal dimension, and predictability metrics, have been widely used to characterize changes in brain dynamics across cognitive states and neuropsychiatric conditions.[23] Therefore, these measures offer insight into how to evaluate human performance in dynamic and high-stress environments.

The complexity synchronization (CS) analysis is a tool for understanding the intercommunications within networks of organ networks, such as within the levels of the central nervous system and between the brain, heart, and lungs.[24] CS is the synchronization in the statistical structure of interacting systems rather than the phase or amplitude. CS is hypothesized as a mechanism for efficient information exchange among organ networks of varying architectures and timescales.[24] Disruptions in synchrony, whether due to cognitive load, fatigue, or degradation of health, can indicate underlying system inefficiencies.[25] Because CS is dependent upon the temporal fluctuations in complexity of physiological time series, we investigated how signal preprocessing impacts both the estimation of complexity and the temporal relationship between signals. However, like complexity, it remains unclear whether observed changes in CS after artifact cleaning represent genuine neural phenomena or artifacts of preprocessing.

The interaction among EEG signal preprocessing, neural complexity, and CS remains an underexplored area. Although artifact removal pipelines aim to enhance signal clarity, their effects on higher-order neural measures are still not fully characterized. There are many different approaches for estimating system complexity and synchronization; however, in this work, we focus on MDEA based on crucial event theory.[26] We investigated how estimates of neural complexity and CS were affected by various artifact removal processes to provide deeper insights into their robustness and reliability. This understanding is essential for refining EEG preprocessing methodologies and improving the interpretability of neural measures in real-world applications.

2. Objectives

This investigation aimed to clarify how artifact removal affected EEG-derived neural complexity and its interpretation across experimental conditions. Specifically, we focused on the following objectives:

- **Objective 1:** Evaluate how neural complexity estimates changed through stages of EEG preprocessing and artifact removal.
- **Objective 2:** Evaluate how CS changed through stages of EEG preprocessing and artifact removal.

By addressing these objectives, this study aimed to improve the reliability of EEG-based insights into brain dynamics and human performance.

3. Methods

3.1 Datasets

For this study, we used two existing datasets that represented different types of artifacts. However, both datasets likely contained persistent line noise at 60 Hz and respiration and cardiac rhythms, which are common in EEG recordings. For Dataset 1, which is dominated by a persistent artifact, 128-channel EEG was recorded at 512 Hz (BioSemi ActiveTwo) while novice users walked with bilateral, ankle exoskeleton assistance during pre- and post-adaptation. Users walked on a treadmill continuously for 5 min before (Pre) and after (Post) a training period where they practiced walking with the exoskeleton assistance for approximately 45 min. In this analysis, we included data from 15 subjects.[27] EEG data collected during walking has been shown to be contaminated with ongoing muscle activity and electrode movement- or disturbance-related artifacts.[28] Dataset 1 also contained transient artifacts such as eye blinks and bursts of increased muscle activity. For Dataset 2, the transient artifact dataset, 64-channel EEG was collected at 2048 Hz (BioSemi ActiveTwo) during a seated virtual-reality shooting task under high- and low-stress conditions.[29] We used four subjects' data, and each condition was approximately 8–10 min long. Subjects were seated and presented with pop-up targets on a simulated range and made movements to orient to the target location. This task produced more transient artifacts related to target presentation with periods of little artifact in-between targets. Transient artifacts include eye blinks, muscle activity, and electrode disturbances associated with head and upper body movements. For Dataset 2, two-lead electrocardiogram (ECG) data were recorded simultaneously with the EEG data using the same data recording system.

3.2 EEG Preprocessing Pipeline

For both datasets, we applied a preprocessing pipeline that has been well characterized for removing artifacts from EEG data during whole body movements.[27] The preprocessing steps were implemented using EEGLAB functions integrated into custom MATLAB scripts. The pipeline included the following stages: 1) 2-Hz, high-pass filter; 2) robust re-referencing, detection, and removal of bad channels[30]; 3) ASR[31]; 4) canonical correlation analysis[32]; and 5) AMICA.[33,34] The order of the artifact removal steps has been tested and is purposeful.[10,27,30,35,36] An important first step is to re-reference the data, because data collected with the BioSemi ActiveTwo is unreferenced and referencing the data alone can drastically improve the common mode rejection ratio.[30] In the initial preprocessing steps, bad channels are also removed, because these channels likely contain little or no brain signal and may have a poor connection with the scalp. The preprocessing pipeline is depicted in Figure 1 along with the EEGLAB functions used to perform each step.

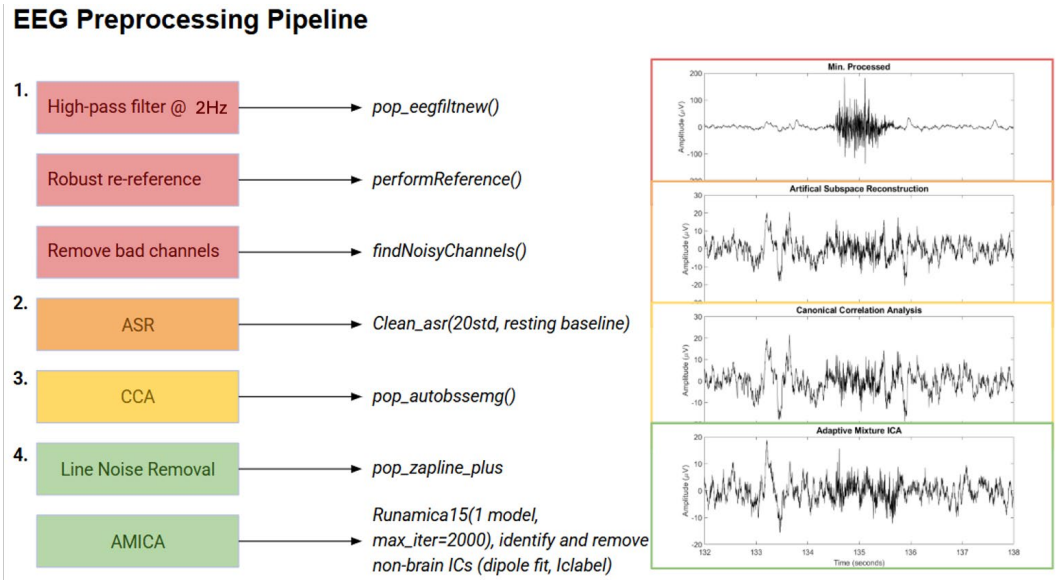


Figure 1. EEG preprocessing pipeline. The left panel shows the sequential preprocessing steps and corresponding EEGLAB/MATLAB functions used in the pipeline. The right panel shows an example of the same EEG signal segment after each successive cleaning stage, illustrating how the signal changes across preprocessing steps.

3.2.1 Minimal Preprocessing

We applied a 2-Hz, high-pass filter to eliminate slow drifts and reduce large walking-related artifacts at the stepping frequency.[37] Next, we performed robust referencing, which estimates a clean reference signal to reduce noise contributions common to all electrodes, and removed bad channels that likely had a poor connection with the scalp.[30]

3.2.2 Artifact Subspace Reconstruction

ASR addresses transient and high-amplitude artifacts caused by eye blinks, muscle activity, or sudden movements.[38] A clean reference dataset, typically from a static baseline condition or a clean section of the dataset, serves as a baseline. ASR detects segments exceeding a standard deviation threshold and reconstructs these portions. We used a standard deviation parameter of 20 to preserve neural variability while attenuating artifacts. We used a baseline collected during quiet standing as the reference for Dataset 1-Persistent Artifact. For Dataset 2-Transient Artifact, we set the ASR function to search for clean segments of data during the condition of interest as the reference

3.2.3 Canonical Correlation Analysis

CCA is a blind source separation technique that identifies highly correlated components across EEG channels.[39] Neural signals are highly autocorrelated, and thus the components with the lowest autocorrelation are identified and removed from the dataset. CCA is particularly effective at targeting muscle activity artifacts because they have a very low autocorrelation.

3.2.4 AMICA with IC Pruning (Maximally Processed)

We applied a line-noise removal procedure (ZapLine+) to suppress 60-Hz power-line interference and its harmonics, and then AMICA was applied. AMICA decomposes EEG signals into statistically ICs using a flexible probabilistic model.[38] AMICA has been shown to be a powerful tool for isolating neural source activity from artifacts. After decomposition into ICs, we fit equivalent current dipoles for each IC using the DIPFIT function in EEGLAB. DIPFIT estimates the location and orientation of a dipole for each IC by finding the dipole that best matches the IC scalp map. True brain sources are theoretically dipolar, and thus we use the residual variance (RV) associated with the dipole fit to help identify brain ICs. The RV reflects the percentage of the IC scalp-map variance that is not explained by the fitted dipole; therefore, lower RV values indicate a better dipole fit, whereas higher RV values indicate a poorer fit. We identify and retain only components that likely represent brain source activity based on the criteria RV

(RV <15%) and ICLabel brain probability (>50%). ICLabel is an automated EEGLAB plugin for classifying ICA components, which was developed as an ML-based classifier that assigns each IC probabilities across seven categories: brain, muscle, eye, heart, line noise, channel noise, and other.[40] In this study, components were retained only when the ICLabel probability for the brain class exceeded 50%. Non-brain components, such as eye or muscle artifacts, were excluded, and the cleaned EEG channel signals were reconstructed for downstream analysis.

3.3 Modified Diffusion Entropy Analysis

MDEA is a statistical approach designed to quantify the complexity of time-series data, such as EEG signals, by examining how temporal fluctuations evolve over time.[41] The theory of self-organized temporal criticality assumes that complex networks function near a critical state (between order and disorder). At this state, they produce "crucial events" (CEs), allowing for adaptive, efficient interactions without central control. Unlike traditional entropy measures, MDEA begins with extracting CEs directly from the raw signal before performing diffusion entropy analysis. This preprocessing step ensures that only meaningful signal transitions contribute to the entropy calculation, reducing the influence of random noise and irrelevant variability.

The MDEA process starts by normalizing the raw EEG signal to a range [0,1], ensuring consistency across datasets. After normalization, the signal is segmented into horizontal *stripes* of optimized size (ΔS).[24] A *CE* is defined whenever the signal crosses from one stripe to another, producing a discrete event marker. Mathematically, these transitions can be represented as follows:

$$\Delta s(t) = s(t) - s(t - \Delta t), \tag{1}$$

where $s(t)$ represents the stripe occupied by the normalized EEG signal at time t, and Δt represents the sampling interval. A nonzero value of $\Delta s(t)$ indicates that the signal has crossed from one stripe to another. Events are then extracted as unit amplitude pulses whenever a stripe boundary is crossed. If the signal remains within the same stripe, no event is recorded, effectively ignoring stationary periods in the signal.

The extracted events are then cumulatively summed to construct a *diffusion trajectory*, which captures the statistical evolution of signal fluctuations over time. The probability distribution function (PDF) of the diffusion trajectory is analyzed to determine how it spreads as a function of the window size w. The diffusion entropy $S(w)$ is calculated as follows:

$$S(w) = -\int_{-\infty}^{\infty} P(x,w) ln[P(x,w)] dx \quad (2)$$

where *P(x,w)* represents the PDF of the diffusion trajectory position *x* for a given window size *w*. CEs play a significant role here, because they represent moments of signal change that are preserved in the diffusion process rather than being diluted by fluctuations that do not produce stripe crossings.

The change in diffusion entropy relative to a reference state is then defined as follows:

$$\Delta S(w) \equiv S(w) - S_{ref} = \delta \ln(w), \quad (3)$$

where $\Delta S(w)$ is the change in diffusion entropy, S_{ref} is the reference entropy, and δ is the scaling index. The scaling index δ is obtained from the slope of $\Delta S(w)$ versus $\ln(w)$, providing a measure of the temporal complexity of the signal.

By focusing on these meaningful events, MDEA can detect patterns of self-organization and critical transitions within the EEG signal. Therefore, CEs are not random but structured temporal markers indicating significant neural dynamics. This extraction and analysis process enables MDEA to robustly distinguish between stochastic noise, structured patterns, and transient critical events, offering a refined measure of neural complexity that remains sensitive to both short-term variability and long-term dependencies

To compute temporal fluctuations in complexity (scaling indices [δ]), we applied MDEA to the EEG and ECG time series data in 30-s moving windows with a 2/3 overlap. The 30-s window length and 2/3 overlap were held constant across all preprocessing stages to ensure direct comparability of the resulting complexity time series. To compute mean complexity for a condition, we averaged all available windows for that condition. We used an adaptive method for MDEA, optimizing stripe size and diffusion trajectory linear region (where the slope is fitted) on every window.[24] We applied MDEA after every successive stage of the EEG preprocessing pipeline to evaluate how each step influenced neural complexity estimates. MDEA was performed sequentially after minimal preprocessing (high-pass filter, re-referencing, bad channel identification, and removal), and then after ASR, followed by CCA, and finally after AMICA pruning. This approach allowed us to track how successive preprocessing stages affected neural complexity estimates. MDEA was applied to each EEG channel, and the mean of all channels was taken for comparisons. We performed a Spearman correlation between the mean EEG complexity time series after each processing

stage to determine whether temporal patterns of complexity were affected by preprocessing.

3.4 Complexity Synchronization

CS was quantified as the Spearman rank correlation between two complexity time series. A rho value of 1 indicates perfect CS, whereas a rho value of 0 indicates no CS. After applying MDEA in 30-s windows with 2/3 overlap, a scaling time series was obtained for each remaining EEG channel at each preprocessing stage. These scaling time series were then averaged across channels to obtain one mean EEG complexity time series for each condition. CS was computed for Dataset 2 between the EEG and ECG complexity time series to assess synchronization between brain and cardiac activity after each preprocessing stage. CS provided a measure of whether preprocessing influenced the relationship in the temporal fluctuations of complexity across systems, beyond its effect on absolute complexity values.

4. Results

4.1 Persistent Artifact Study

We found that in data with persistent artifacts, temporal fluctuations are preserved and mean complexity significantly decreases from minimally processed data to maximally processed data for both Pre- and Post-adaptation experimental conditions ($p = 0.0101$ and $p = 0.0173$, one-sided Wilcoxon p-value, respectively). There is a significant decrease in standard deviation from minimally to maximally processed data for pre and post ($p = 0.0368$ and $p = 0.0173$, one-sided Wilcoxon p-value, respectively). Figure 2a shows an example of one subject in Pre-adaptation for each compounded step. We found that the EEG scaling indices time series had strong, positive correlations after each preprocessing step. Figure 2b is a Fisher-correct Spearman rank correlation heat matrix annotating the Spearman rho between all EEG processing pipeline steps in Pre- (blue) and Post- (red) adaptation averaged across subjects.

Figure 2c reveals that through all cleaning steps, the relationship between conditions (i.e., Pre and Post) is preserved with an increase in complexity from Pre- to Post-adaptation for minimally processed, ASR, CCA, and maximally processed data ($p = 0.0153$, $p = 0.0224$, $p = 0.0198$, and $p = 0.0028$, one-sided Wilcoxon p-value, respectively).

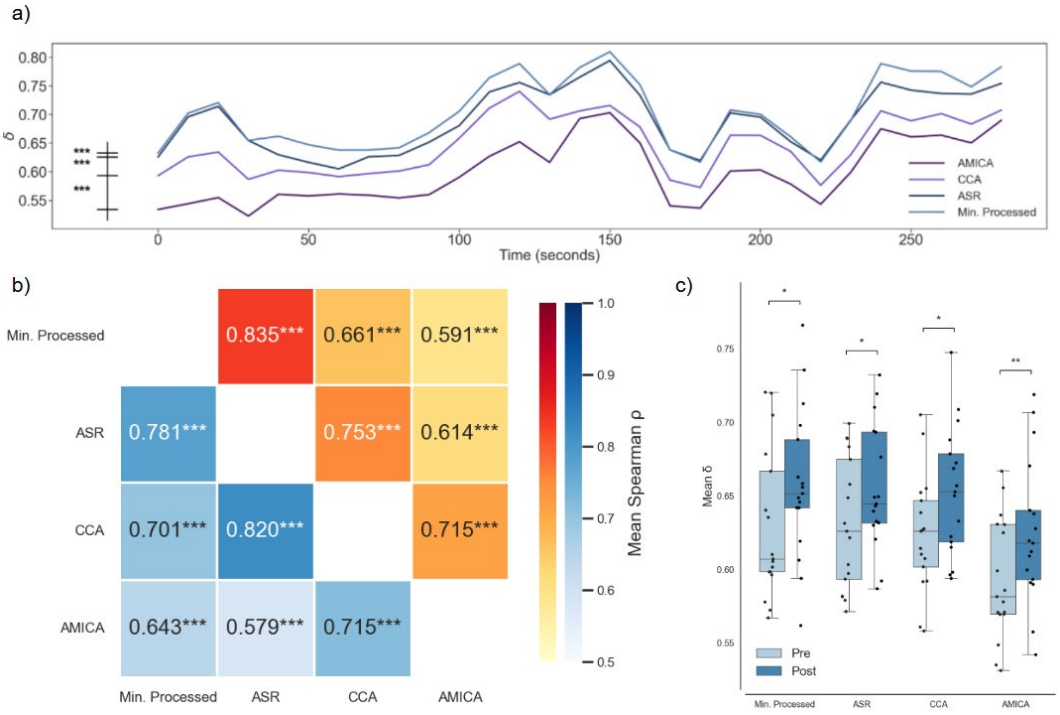


Figure 2. MDEA on data with persistent artifacts. a) An example of one subject for one condition where temporal fluctuations are preserved through all processing steps in Pre-adaptation. b) A Spearman rank correlation heat matrix annotating the Fisher-z corrected averaged Spearman rho between all EEG processing pipeline steps in Pre- (blue) and Post- (red) adaptation averaged across subjects. c) A bar plot comparing the mean complexities (δ) across the EEG cleaning pipeline over time. Statistical significance is denoted by an asterisk: *p < 0.05, **p < 0.01, and ***p < 0.001.

4.2 Transient Artifact Study

We found that in data with primarily transient artifacts, temporal fluctuations were preserved, and the mean complexity from minimally processed data to maximally processed data decreased, albeit not significantly, which is likely due to the small sample size ($p = 0.0625$, one-sided Wilcoxon p-value) as well as a decrease in standard deviation from minimally to maximally processed data ($p = 0.0625$, one-sided Wilcoxon p-value). Figure 3a shows an example of one subject in high stress for each preprocessing stage. The bracketed asterisk indicates a significant drop in mean scaling after each step in the processing pipeline. Figure 3b demonstrates that the temporal fluctuations in complexity of the EEG signals are largely preserved in instances in which all correlations between preprocessing stages have moderate to high values. Figure 3c depicts mean complexity estimates for each condition (high and low stress) for each preprocessing stage. We did not find a significant difference in high- and low-stress conditions for any preprocessing stage. Figure 3d evaluates the CS between EEG and ECG after each preprocessing stage. For low and high stress, CS remained significant before and after preprocessing.

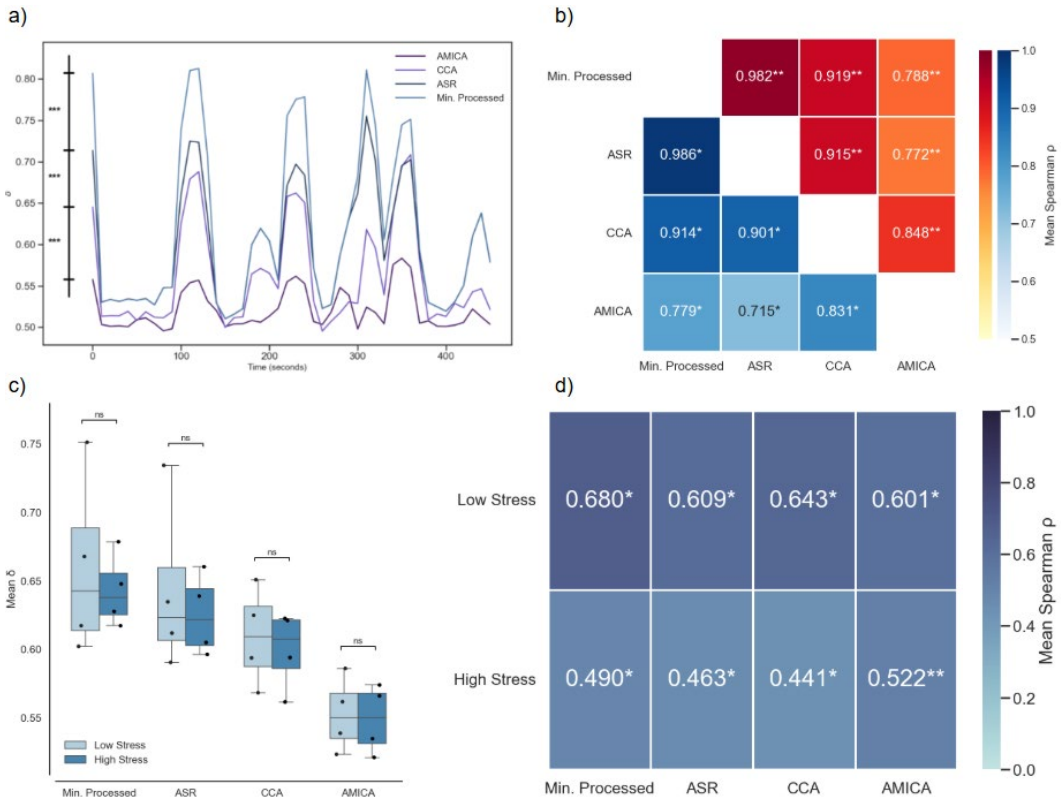


Figure 3. MDEA on data with spurious artifacts. a) An example of one subject whose temporal fluctuations are preserved through all processing steps in high stress, with an asterisk denoting significant decrease in mean scaling. b) A Spearman rank correlation heat matrix annotating the Fisher-z corrected averaged Spearman rho between all EEG processing pipeline steps in low (blue) and high (red) stress averaged across subjects. c) A bar plot comparing the mean complexities (δ) across the EEG cleaning pipeline over time. d) A Spearman correlation heat matrix annotating the Fisher-z corrected averaged Spearman rho between EEG and ECG scaling indices across each cleaning step in the processing pipeline. Statistical significance is denoted by an asterisk: *p < 0.05 and **p < 0.001.

5. Discussion

In this study, we aimed to understand how EEG preprocessing and artifact removal affected complexity and CS as estimated by MDEA. We found decreased mean complexity as we progressed through the preprocessing pipeline. However, temporal fluctuations in complexity of the neural signals were preserved throughout the preprocessing pipeline stages for both transient and persistent artifact-dominated datasets. The magnitude of MDEA complexity estimates should be interpreted with caution with regards to the presence of artifacts and any artifact removal techniques applied to the data. Temporal fluctuations in complexity estimated by MDEA were robust to artifacts and our preprocessing pipeline.

5.1 Reduction in Neural Complexity and Variability with Artifact Removal

We observed a decrease in mean and standard deviation complexity estimates across both transient and persistent artifact datasets when comparing minimally to maximally processed EEG data. The ASR and CCA preprocessing steps primarily target movement artifacts, eye blink- and eye movement-related artifacts, and muscle activity artifacts.[31,32] AMICA attempts to isolate neural activity from the remaining artifacts.[33,34] Based on those assumptions, we interpret the reduction in complexity as a result of the reduction of artifacts in the recorded EEG data. The reduction in complexity is not surprising given that each step of the pipeline removes information or variability from the signal by separating or attenuating artifacts. The concurrent decrease in standard deviation of the group mean complexity is another indication of reduced noise in the signal. Our main takeaway from these findings is that the magnitude of complexity of EEG signals should be interpreted carefully because persistent and transient artifact contamination can overinflate complexity estimates.

Although we did not directly compare MDEA-derived complexity with conventional oscillatory power metrics, the present findings raise the possibility that complexity estimates may be less sensitive to certain artifact types. Oscillatory power measures are known to be strongly influenced by physiological artifacts. For example, muscle activity produces broadband signals that overlap neural frequency bands and can substantially inflate high-frequency EEG power measurements.[42] In contrast, the preservation of condition-level relationships across preprocessing stages in the persistent-artifact dataset suggests that complexity measures may retain meaningful signal structure even when artifacts are present. Although this observation does not demonstrate that complexity metrics are inherently more robust than oscillatory measures, it highlights the potential for complexity-based analyses to capture neural dynamics despite the presence of artifacts.

Although the reduction in complexity is primarily due to a reduction of artifacts in the EEG data, it is possible that the preprocessing pipelines are distorting the neural signals. However, our results suggest that the ability to detect relative changes in neural complexity between conditions was not affected by our preprocessing pipeline. One demonstration is the result of comparing mean complexity during Pre- and Post-adaptation while walking with the exoskeleton (Figure 2). In both conditions, the subjects were walking for 5 min continuously on a treadmill while receiving assistance from an exoskeleton. Artifacts during these conditions were similar, given that the tasks of walking at a constant speed were consistent across Pre- and Post-adaptation phases[27]; however, the relationship of decreased

complexity during late adaptation is maintained throughout the preprocessing pipeline. In the same dataset, Song et al.[27] found increased alpha band power in sensorimotor brain regions during Post-adaptation, which was interpreted as decreased cortical activity. Thus, decreased complexity may indicate decreased sensorimotor neural activity during Post adaptation. Furthermore, the finding of decreased complexity after habituation agrees with the results from Hua et al.[43] who found a decrease in sensorimotor EEG complexity (fractal dimension analysis) after habituation to a standing balance task.

In both transient and persistent artifact datasets, the temporal fluctuations of complexity were preserved across preprocessing steps. A preservation of temporal fluctuations after artifact removal indicates that complexity over time is primarily driven by the neural signal and not artifact. For the transient dataset, we were able to observe CS between heterogeneous systems, the brain and heart. The synchronization is characterized by matching multifractal dimensions in their respective time series, which suggests efficient communications and optimal physiological function. The CS between the averaged EEG and ECG is significant after each stage of cleaning in low and high stress, but it is consistently higher in the low-stress condition. Therefore, although temporal fluctuations of EEG are largely preserved across cleaning steps, to accurately assess performance of heterogeneous-coupled systems, EEG preprocessing may be necessary for an accurate understanding of the synchrony between these two systems.

The preservation of the relationships between conditions in persistent and transient artifacts combined with the preservation of temporal fluctuations suggest that fewer EEG preprocessing steps may be necessary. EEG preprocessing can require a non-trivial amount of time and computational power depending on the dimensionality of the data. ICA, a common preprocessing step, takes a notoriously long time to process due to the high dimensionality and large size of EEG datasets, which require extensive matrix computations and iterative optimization.[39] Reducing EEG preprocessing requirements also reduces computational strain, allowing resources to be reallocated for more detailed downstream analyses or supporting larger datasets. For smaller research teams with limited access to high-performance computing resources, shorter runtimes decrease the challenges to conducting large-scale EEG studies. Furthermore, faster EEG preprocessing has practical value in translating complexity-based neural monitoring into brain computer interfaces. In wearable exoskeletons, for example, reduced processing latency could enable real-time updates to control algorithms, incorporating neural complexity metrics into multi-cost optimization functions without significant lag.[44]

EEG recordings are highly susceptible to contamination from physiological and environmental artifacts such as eye movements, muscle activity, cardiac signals,

and electrical interference. As a result, artifact detection and removal are widely considered essential preprocessing steps in EEG analysis pipelines[11,18] Previous work has also shown that preprocessing pipelines can reduce artifact-related variability and improve the reliability of downstream EEG analyses, including measures such as functional connectivity.[38,39] In this context, the preservation of relationships between conditions and temporal fluctuations observed in the present study suggests that some signal features may remain detectable despite artifact contamination, potentially reducing the need for extensive preprocessing in certain applications.

5.2 Neural and Organ System Synchrony and CS

Research has shown that synchrony is closely tied to critical dynamics within the brain, such as phase transitions and state changes in neural networks.[45–47] These dynamics are often linked to *CEs*, where transient increases in synchrony correspond to significant cognitive or physiological states.[48–50] CS facilitates efficient information exchange across brain and organ networks, enabling the integration of information and coordination of neural processes essential for adaptive behavior. This dynamic interplay is not just a result of signal processing, but it also reflects the brain's and body's underlying capacity for resilience and responsiveness in the face of varying cognitive and environmental demands.

Despite the aggressive artifact removal steps used in preprocessing, heart–brain CS was maintained across each EEG preprocessing stage. This observation suggests that the preprocessing pipeline is not simply amplifying artificial synchrony, but it is effectively preserving biologically relevant neural dynamics. The ability to retain both CS and temporal variability through successive cleaning stages strengthens confidence in the integrity of our EEG processing pipeline and the complexity metric, reinforcing its validity for studying system-level performance, cognitive adaptation, and brain-environment interactions.

6. Limitations

One limitation of this study is that we do not know the actual statistical structure of the neural signal. Thus, we cannot determine exactly how artifacts and the subsequent preprocessing analysis affect the neural signal. Future studies may add to the understanding of how artifacts and their removal impact estimates of complexity by using 1) synthetic data or 2) real data with simulated artifacts. Furthermore, the studies used in this investigation were not conducted with the intent of using MDEA to examine neural complexity. The experimental paradigms were not designed to specifically manipulate neural complexity or CS. Although

the results discussed are promising, future investigations should consider both systematically manipulating neural complexity and the presence of artifacts. This knowledge would be useful for insights into real-world brain function, because those scenarios often induce large amounts of artifact and dynamic changes in brain state.

Window length and overlap may influence the temporal resolution and statistical stability of MDEA-derived complexity time series and, consequently, estimates of CS. Because these parameters were held fixed in the present analysis, we did not assess the sensitivity of the observed preservation of temporal fluctuations and CS to alternative windowing parameters. Future work should evaluate the robustness of these findings across different window lengths and overlap values.

7. Conclusion

In this investigation, we explored the impact of artifact removal techniques on EEG-derived estimates of neural complexity across datasets. By applying MDEA at cumulative stages of preprocessing—minimal processing, ASR, CCA, and AMICA with ICA pruning—we observed a consistent preservation of temporal fluctuations in neural complexity estimates despite successive artifact removal steps. The results demonstrated that although artifact cleaning reduced mean overall complexity, the relationship between experimental conditions remained intact. These insights provide clarity on the nuanced interplay among preprocessing pipelines, neural complexity, and synchrony, advancing the interpretability of EEG-derived metrics in both research and applied settings. Future research should focus on designing EEG experiments for neural complexity analysis, minimizing confounding factors and ensuring controlled artifact characteristics. Investigations across diverse experimental paradigms, including prolonged locomotion tasks, cognitive workload assessments, and dynamic human-system interaction scenarios, will further validate the robustness and applicability of MDEA-derived metrics.

8. References

1. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol. 2000;278(6):H2039–H2049. https://doi.org/10.1152/ajpheart.2000.278.6.H2039

2. Tononi G, Edelman GM, Sporns O. Complexity and coherency: integrating information in the brain. Trends Cogn Sci. 1998;2(12):474–484. https://doi.org/10.1016/s1364-6613(98)01259-5.

3. McIntosh AR, Kovacevic N, Itier RJ. Increased brain signal variability accompanies lower behavioral variability in development. PLoS Comput Biol. 2008;4(7):e1000106. https://doi.org/10.1371/journal.pcbi.1000106

4. Vakorin V, Lippé S, McIntosh A. Variability of brain signals processed locally transforms into higher connectivity with brain development. J Neurosci. 2011;31(17):6405–6413. https://doi.org/10.1523/jneurosci.3153-10.2011

5. Urigüen JA, Garcia-Zapirain B. EEG artifact removal-state-of-the-art and guidelines. J Neural Eng. 2015;12(3):031001. https://doi.org/10.1088/1741-2560/12/3/031001

6. Rubinos C, Alkhachroum A, Der-Nigoghossian C, Claassen J. Electroencephalogram monitoring in critical care. Semin Neurol. 2020;40(6):675–680. https://doi.org/10.1055/s-0040-1719073

7. Diachenko M et al. Improved manual annotation of EEG signals through convolutional neural network guidance. eNeuro. 2022;9(5):ENEURO.0160-22.2022. https://doi.org/10.1523/eneuro.0160-22.2022

8. Kappel SL, Looney D, Mandic DP, Kidmose P. Physiological artifacts in scalp EEG and ear-EEG. Biomed Eng Online. 2017;16(1):103. https://doi.org/10.1186/s12938-017-0391-2

9. Gorjan D, Gramann K, De Pauw K, Marusic U. Removal of movement-induced EEG artifacts: current state of the art and guidelines. J Neural Eng. 2022;19(1):ac542c. https://doi.org/10.1088/1741-2552/ac542c

10. Snyder KL, Kline JE, Huang HJ, Ferris DP. Independent component analysis of gait-related movement artifact recorded using EEG electrodes during treadmill walking. Front Hum Neurosci. 2015;9:639. https://doi.org/10.3389/fnhum.2015.00639

11. Jiang X, Bian GB, Tian Z. Removal of artifacts from EEG signals: a review. Sensors (Basel). 2019;19(5):987. https://doi.org/10.3390/s19050987

12. James CJ, Hesse CW. Independent component analysis for biomedical signals. Physiol Meas. 2005;26(1):R15–R39. https://doi.org/10.1088/0967-3334/26/1/R02

13. Blum S, Jacobsen NSJ, Bleichner MG, Debener S. A Riemannian modification of artifact subspace reconstruction for EEG artifact handling. Front Hum Neurosci. 2019;13:141. https://doi.org/10.3389/fnhum.2019.00141

14. Lin CT et al. Real-time EEG signal enhancement using canonical correlation analysis and Gaussian mixture clustering. J Healthc Eng. 2018;2018:5081258. https://doi.org/10.1155/2018/5081258

15. Dai Y et al. A fast approach to removing muscle artifacts for EEG with signal serialization based ensemble empirical mode decomposition. Entropy (Basel). 2021;23(9):1170. https://doi.org/10.3390/e23091170

16. Zeng K et al. An EEMD-ICA approach to enhancing artifact rejection for noisy multivariate neural data. IEEE Trans Neural Syst Rehabil Eng. 2016;24(6):630–638. https://doi.org/10.1109/tnsre.2015.2496334

17. Hsu SH et al. Modeling brain dynamic state changes with adaptive mixture independent component analysis. Neuroimage. 2018;183:47–61. https://doi.org/10.1016/j.neuroimage.2018.08.001.

18. Labate D, La Foresta F, Mammone N, Carlo F. Effects of artifacts rejection on EEG complexity in Alzheimer's disease. In: Smart Innovation, Systems and Technologies. Vol 37. Springer; 2015. p. 129–136. https://doi.org/10.1007/978-3-319-18164-6_13

19. Werner G. Consciousness related neural events viewed as brain state space transitions. Cogn Neurodyn. 2009;3(1):83–95. https://doi.org/10.1007/s11571-008-9040-6

20. Chiel HJ, Beer RD. The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. Trends Neurosci. 1997;20(12):553–557. https://doi.org/10.1016/s0166-2236(97)01149-1.

21. Wan W et al. Resting state EEG complexity as a predictor of cognitive performance. Physica A. 2023;624:128952. https://doi.org/10.1016/j.physa.2023.128952

22. Zhai Y et al. Mental fatigue decreases complexity: evidence from multiscale entropy analysis of instantaneous frequency variation in alpha rhythm. Front Hum Neurosci. 2022;16:906735. https://doi.org/10.3389/fnhum.2022.906735.

23. Lau ZJ, Pham T, Chen SHA, Makowski D. Brain entropy, fractal dimensions and predictability: a review of complexity measures for EEG in healthy and neuropsychiatric populations. Eur J Neurosci. 2022;56(7):5047–5069. https://doi.org/10.1111/ejn.15804

24. Schizas I et al. Complexity synchronization analysis of neurophysiological data: theory and methods. Front Netw Physiol. 2025;5:1570530. https://doi.org/10.3389/fnetp.2025.1570530.

25. Lewis MD. Bridging emotion theory and neurobiology through dynamic systems modeling. Behav Brain Sci. 2005;28(2):169–245. https://doi.org/10.1017/s0140525x0500004x.

26. West B, Geneston E, Grigolini P. Maximizing information exchange between complex networks. Phys Rep. 2008;468(1–3):1–99. https://doi.org/10.1016/j.physrep.2008.06.003

27. Song S, Haynes CA, Bradford JC. Novice users' brain activity and biomechanics change after practicing walking with an ankle exoskeleton. IEEE Trans Neural Syst Rehabil Eng. 2024;32:2522–2532.

28. Richer N, Bradford JC, Ferris DP. Mobile neuroimaging: what we have learned about the neural control of human walking, with an emphasis on EEG-based research. Neurosci Biobehav Rev. 2024;162:105718. https://doi.org/10.1016/j.neubiorev.2024.105718

29. Kerick SE et al. Neural and behavioral adaptations to frontal theta neurofeedback training: a proof of concept study. PLoS One. 2023;18(3):e0283418. https://doi.org/10.1371/journal.pone.0283418

30. Bigdely-Shamlo N, Mullen T, Kothe C, Su KM, Robbins KA. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. Front Neuroinform. 2015;9:16. https://doi.org/10.3389/fninf.2015.00016

31. Chang CY, Hsu SH, Pion-Tonachini L, Jung TP. Evaluation of artifact subspace reconstruction for automatic EEG artifact removal. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018 July 18–21; Honolulu, HI. IEEE; c2018. p. 1242–1245. https://doi.org/10.1109/embc.2018.8512547

32. Safieddine D et al. Removal of muscle artifact from EEG data: comparison between stochastic (ICA and CCA) and deterministic (EMD and Wavelet-based) approaches. EURASIP J Adv Signal Process. 2012;2012(1):127. https://doi.org/10.1186/1687-6180-2012-127

33. Frank G et al. An exploration of optimal parameters for efficient blind source separation of EEG recordings using AMICA. In: 2023 IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE); 2023 Dec; Taichung, Taiwan. IEEE; c2023. p. 205–210.

34. Leutheuser H et al. Comparison of the AMICA and the InfoMax algorithm for the reduction of electromyogenic artifacts in EEG data. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2013 July 3–7; Osaka, Japan. IEEE; c2013. p. 6804–6807. https://doi.org/10.1109/EMBC.2013.6611119

35. Nordin AD, Hairston WD, Ferris DP. Faster gait speeds reduce alpha and beta EEG spectral power from human sensorimotor cortex. IEEE Trans Biomed Eng. 2020;67(3):842–853. https://doi.org/10.1109/tbme.2019.2921766

36. Symeonidou ER, Nordin AD, Hairston WD, Ferris DP. Effects of cable sway, electrode surface area, and electrode mass on electroencephalography signal quality during motion. Sensors (Basel). 2018;18(4):1073. https://doi.org/10.3390/s18041073

37. Russell DM, Apatoczky DT. Walking at the preferred stride frequency minimizes muscle activity. Gait Posture. 2016;45:181–186. https://doi.org/10.1016/j.gaitpost.2016.01.027

38. Chang CY, Hsu SH, Pion-Tonachini L, Jung TP. Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel EEG recordings. IEEE Trans Biomed Eng. 2020;67(4):1114–1121. https://doi.org/10.1109/TBME.2019.2930186

39. Li J et al. A unified canonical correlation analysis-based framework for removing gradient artifact in concurrent EEG/fMRI recording and motion artifact in walking recording from EEG signal. Med Biol Eng Comput. 2017;55(9):1669–1681. https://doi.org/10.1007/s11517-017-1620-3

40. Pion-Tonachini L, Kreutz-Delgado K, Makeig S. ICLabel: an automated electroencephalographic independent component classifier, dataset, and website. Neuroimage. 2019;198:181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026

41. Culbreth G, Baxley J, Lambert D. Detecting temporal scaling with modified diffusion entropy analysis. arXiv; 2023 Nov 13. arXiv:2311.11453. https://doi.org/10.48550/arXiv.2311.11453

42. Muthukumaraswamy SD. High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. Front Hum Neurosci. 2013;7:138. https://doi.org/10.3389/fnhum.2013.00138

43. Hua A et al. Nonlinear dynamics of postural control system under visual-vestibular habituation balance practice: evidence from EEG, EMG and center of pressure signals. Front Hum Neurosci. 2024;18:1371648. https://doi.org/10.3389/fnhum.2024.1371648

44. Zhang J et al. Human-in-the-loop optimization of exoskeleton assistance during walking. Science. 2017;356(6344):1280–1284. https://doi.org/10.1126/science.aal5054

45. Kazemi S, Farokhniaee A, Jamali Y. Criticality and partial synchronization analysis in Wilson-Cowan and Jansen-Rit neural mass models. PLoS One. 2024;19(7):e0292910. https://doi:10.1371/journal.pone.0292910

46. Buendía V, Villegas P, Burioni R, Muñoz MA. The broad edge of synchronization: Griffiths effects and collective phenomena in brain networks. Philos Trans A Math Phys Eng Sci. 2022;380(2227):20200424. https://doi.org/10.1098/rsta.2020.0424

47. Wang S et al. Critical-like brain dynamics in a continuum from second- to first-order phase transition. J Neurosci. 2023;43(15):2715–2728. https://doi.org/10.1523/jneurosci.1889-22.2023

48. West BJ, Grigolini P, Kerick SE, Franaszczuk PJ, Mahmoodi K. Complexity synchronization of organ networks. Entropy (Basel). 2023;25(10):1393. https://doi.org/10.3390/e25101393

49. Mahmoodi K, Kerick SE, Grigolini P, Franaszczuk PJ, West BJ. Complexity synchronization: a measure of interaction between the brain, heart and lungs. Sci Rep. 2023;13(1):11433. https://doi.org/10.1038/s41598-023-38622-8

50. Nasrat S et al. Crucial events identify early stage of cardiac autonomic neuropathy progression from ECG signals. In: 2024 46th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2024 July 15–19; Orlando, FL. IEEE; c2024. p. 1–4. https://doi.org/10.1109/EMBC53108.2024.10782197

List of Symbols, Abbreviations, and Acronyms

AMICA	Adaptive Mixture Independent Component Analysis
ASR	Artifact Subspace Reconstruction
CCA	Canonical Correlation Analysis
CE	Crucial Event
CS	Complexity Synchronization
ECG	Electrocardiogram
EEG	Electroencephalogram
IC	Independent Components
ICA	Independent Component Analysis
MDEA	Modified Diffusion Entropy Analysis
ML	Machine Learning
PDF	Probability Distribution Function
RV	Residual Variance

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DEVCOM ARL FCDD RLB CB TECH LIB